

SIMULATIONS OF RNA INTERACTIONS WITH MONOVALENT IONS

Alan A. Chen,^{*,†} Marcelo Marucho,^{*,†} Nathan A. Baker,^{*,‡,†}
and Rohit V. Pappu^{*,†,§,¶}

Contents

1. Introduction	412
2. Finite Size Artifacts in All-Atom Simulations of Ion–Nucleic Acid Interactions	414
3. On the Use of “Neutralizing Counterions Only” in All-Atom Simulations	421
4. Results from Simulations of Canonical A-Form RNA and B-Form DNA Helices	421
5. Comparison with Predictions of the Nonlinear Poisson–Boltzmann Equation	426
Acknowledgment	430
References	430

Abstract

RNA folding and binding reactions are mediated by interactions with ions that make up the surrounding aqueous electrolytic milieu. Although Mg^{2+} ions are often implicated as being crucial for RNA folding, it is known that folding is feasible in high concentrations of monovalent alkali-halide salts. Experiments have yielded important information regarding the salt dependence of RNA stability. Recent work has shown that molecular simulations based on explicit representations of solvent molecules and monovalent ions can also provide useful insights regarding the ionic atmospheres around model RNA systems. These insights can help rationalize intriguing observations regarding the dependence of RNA stability on cation type providing one pays attention to important considerations that go into the proper design of molecular

* Computational and Molecular Biophysics Program, Washington University in St. Louis, St. Louis, Missouri, USA

† Center for Computational Biology, Washington University in St. Louis, St. Louis, Missouri, USA

‡ Department of Biochemistry and Molecular Biophysics, Washington University in St. Louis, St. Louis, Missouri, USA

§ Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri, USA

¶ Washington University in St. Louis, St. Louis, Missouri, USA

simulations. These issues are discussed in detail and the methods are applied to an A-form RNA and B-form DNA sequence. The results of these simulations are compared to previous work on a kissing-loop system with analogous sequence. In particular, ionic atmospheres obtained from molecular simulations are compared to those obtained using the nonlinear Poisson Boltzmann model for continuum electrostatics for these three different nucleic acid systems. The comparisons indicate reasonable agreement in terms of coarse-grained observables such as the numbers of counterions accumulated around the solutes. However, details of the ionic atmospheres, captured in terms of spatial free energy density profiles, are quite different between the two approaches. These comparisons suggest the need for improvements in continuum models to capture sequence-specific effects, ion-ion correlation, and the effects of partial dehydration of ions.

1. INTRODUCTION

The folding and binding reactions of RNA molecules are mediated by mobile counterions in the surrounding milieu (Draper, 2004; Woodson, 2008). These reactions often require millimolar amounts of divalent Mg^{2+} ions (Laing *et al.*, 1994; Pan *et al.*, 1999; Rangan and Woodson, 2003; Takamoto *et al.*, 2004). Monovalent ions such as Na^+/K^+ also play an important role in RNA folding although the required concentrations of monovalent ions are considerably larger than Mg^{2+} ions (Das *et al.*, 2003; Draper, 2004; Perez-Salas *et al.*, 2004; Takamoto *et al.*, 2004).

Several RNA systems are capable of folding/binding in the absence of Mg^{2+} and such systems have been the focus of recent investigations (Lambert *et al.*, 2009). These studies have identified intriguing dependencies of RNA stability on monovalent counterion type that have merited closer scrutiny through computational methods. An elegant example is the Tar–Tar[★] kissing-loop system studied by Draper and coworkers (Lambert *et al.*, 2009). Tar–Tar[★] is composed of a pair of 16 nucleotide hairpins, each containing a 5 bp stem and a six nucleotide loop (Chang and Tinoco, 1997). The two loops have complementary sequences that stabilize the complex via intermolecular Watson–Crick base pairs. This results in substantial deformation of the RNA backbone, with phosphorus atoms from opposing strands (C6 and U22) approaching within 5.25 Å. This distance of approach is small in comparison to the smallest possible interstrand separation of 10.4 Å between phosphorus atoms in canonical A-form duplexes. Lambert *et al.* measured the stability of the Tar–Tar[★] complex in a series of alkali-chloride salts using isothermal titration and thermal melt experiments. The stability of the Tar–Tar[★] complex is inversely proportional to the crystallographic radius of the monovalent counterion. Conversely, folding of the isolated Tar[★] hairpin shows only marginal dependence on

counterion type. [Chen *et al.* \(2009\)](#) used all-atom molecular dynamics simulations with explicit representations of ions and water molecules to quantify differences in the ionic atmospheres of Na^+ , K^+ , and Cs^+ , respectively around the Tar–Tar[★] complex. These simulations yielded insights that appear to rationalize the experimentally observed trends regarding the dependence of Tar–Tar[★] stability on the identity of monovalent counterions. The work of [Chen *et al.*](#) also yielded lessons for the optimal way to design molecular dynamics simulations that are focused on quantitative studies of ionic atmospheres around nucleic acids. These lessons encompass a range of issues including the selection of forcefield parameters for nucleic acids and ions, the choice of system size (numbers of water molecules and ions) for a target salt concentration, the specification of simulation length, and the choice of suitable analysis methods required to obtain detailed ionic atmosphere information. This chapter provides a discussion of these issues and the considerations that must be taken into account when setting up all atom molecular dynamics simulations to study ion interactions with nucleic acids. Following our detailed discussion of these considerations, we present representative results for A-form RNA and B-form DNA sequences and contrast these with results obtained by [Chen *et al.*](#) for the Tar–Tar[★] system. The sequences for the canonical RNA/DNA structures are analogous to that of Tar–Tar[★] ([Fig. 20.1](#)). This allows us to query the effects of three-dimensional structural changes on ionic atmospheres. The results demonstrate the capabilities of modern simulation methods and the complementary role such simulations play to experimental approaches that interrogate the interactions of ions with nucleic acids.

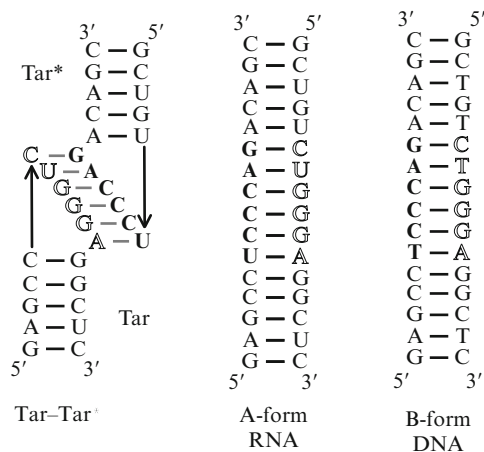


Figure 20.1 Details of the sequences for the Tar–Tar[★], A-form RNA, and B-form DNA systems discussed in this work.

Our analyses of what is needed to obtain meaningful results from all atom simulations with explicit representations of ions and solvent molecules also reveal the limitations of atomistic approaches—specifically, the difficulty of scaling up these simulations to study conformational transitions, binding–unbinding reactions, and the details of salt dependencies of such reactions. Therefore, in [Section 5](#), we provide a comparative characterization of coarser-grained continuum solvation models based mainly on the nonlinear Poisson–Boltzmann equation. This comparison is done with a view toward assessing the suitability and accuracy of fast, novel, albeit empirical implicit solvent models ([Vitalis and Pappu, 2009](#)) for future use in modeling ion-mediated processes for RNA molecules. Based on the encouraging results from our comparative studies, we propose a general approach that incorporates implicit solvent models in calculations with RNA molecules.

2. FINITE SIZE ARTIFACTS IN ALL-ATOM SIMULATIONS OF ION–NUCLEIC ACID INTERACTIONS

Molecular dynamics simulations attempt to represent a macroscopic system using a combination of periodic boundary conditions and small, fixed numbers of solvent molecules in the central simulation cell. In these simulations of solvent plus a single macromolecular solute, the sizes of simulation cells determine the effective solute concentration. Hence, the size of the simulation cell needs to be chosen to ensure that interactions between images of macromolecules are minimal and the simulation mimics a dilute macromolecular solution. This is usually the first consideration in setting up the simulation. For systems with dipolar macromolecular solutes and solvents (such as water), the size of the central simulation cell determines whether the bath contains sufficient numbers of solvent molecules to represent the fraction of solvent molecules that are perturbed by the presence of the macromolecule and the fraction of molecules that display bulk-like behavior. Typical molecular dynamics simulations are run for hundreds of nanoseconds. The larger the simulation cell, the larger the number of force evaluations per time step, and hence the longer the “wall clock” time needed to complete the simulations. Given the computational cost of energy and force evaluations in molecular simulations, there is considerable motivation to use the smallest possible simulation cells to make simulations practically feasible. When the solvent is dipolar, for example water, the perturbation of solvent structure by a low-charge density macromolecule is short-range in nature and does not exceed two to three layers of solvent molecules ([Smolin and Winter, 2004](#)). Therefore, increasing box sizes beyond minimal requirements is not necessary, and in fact additional solvent molecules will not have a dominant effect on the behavior of the macromolecule.

The preceding discussion focused on considerations that go into choosing parameters for simulation cell sizes and numbers of solvent molecules when the systems comprise of low-charge density solutes and dipolar solvents. This discussion also applies, with appropriate modifications, to systems with solutes that have small numbers of charged moieties when compared to the number of dipolar entities. However, these considerations change when considering the setup for simulating systems that include highly charged solutes such as RNA or DNA. Such systems are characterized by significant intramolecular and solvent-mediated long-range electrostatic interactions and the solvent must include mono- and multivalent counter- and coions. Most methods for handling long-range electrostatic interactions require electroneutrality within the central simulation cell; this requirement sets the minimal number of counterions to be included in the simulation.

For macromolecules such as RNA and DNA, there is a significant accumulation of counterions and a concomitant depletion of coions within a local volume around the macroion (Anderson and Record, 1995). Proper recovery of the partitioning between local and bulk ions can only occur if sufficiently large numbers of ions are present in the simulation. Recent work focused on the simulation of ionic atmospheres of monovalent alkali halide salts around a model 32-nucleotide kissing-loop system known as Tar-Tar[★] (Chen *et al.*, 2009). The target salt concentrations for these simulations were 800 mM and the goal was to provide an explanation for the experimentally observed cation-specific stability of the Tar-Tar[★] system. Systematic investigations were performed to identify the minimal setup required to achieve satisfactory, if less than perfect partitioning between local and bulk ionic environments. Such a setup consisted of ~ 120 Å cubic boxes encompassing $\sim 55,000$ water molecules, ~ 800 ion pairs, and 30 neutralizing counterions. Similar parameters ($\sim 10^3$ excess ion pairs and $\sim 6 \times 10^4$ water molecules) were found to be necessary in comparative simulations of 16 bp RNA or DNA duplexes. The results of these simulations will be described in the following sections. The sizes of these simulation cells are substantially larger than the molecular dimensions of the RNA/DNA solutes, which are approximate cylinders with radii of ~ 15 Å and long axes of ~ 50 Å. We found that smaller simulation cells with $O(10^2)$ excess ion pairs result in incorrect partitioning between local and bulk ionic environments. Such inaccuracies can be detected using methods that are described below.

The most easily diagnosed finite size artifact is the presence of a concentration mismatch between the counter- and coion to water ratio (molality) and the actual electrolyte concentration at the “edge” of the simulation box. This is detected by calculating the macromolecule-counterion and macromolecule-coion radial distribution functions (RDFs) from an equilibrated trajectory. Examples of RDFs based on previous work on the

Tar–Tar[★] system (Chen *et al.*, 2009) are shown in Fig. 20.2. Panels A and B of Fig. 20.2 show the RNA phosphate-counter-/coion RDFs using simulation data based on 80 and 120 Å simulation cells, each containing 236 and 792 excess ion pairs, respectively. In the 120 Å box, both RDFs decay to unity (within 1%) as the edge of the central simulation cell is approached. This indicates the presence of a clear local-bulk partition and the presence of a bulk-like reservoir with which ions can be exchanged. The phosphate-counterion and phosphate-coion RDFs differ by a few percent in this regime, but these differences are due to the fact that there are always 30 additional neutralizing counterions present to ensure net electroneutrality of the Tar–Tar[★] system. Conversely, at the boundary of the 80 Å box, the coion density exceeds unity by $\sim 15\%$. This artifact is the result of the

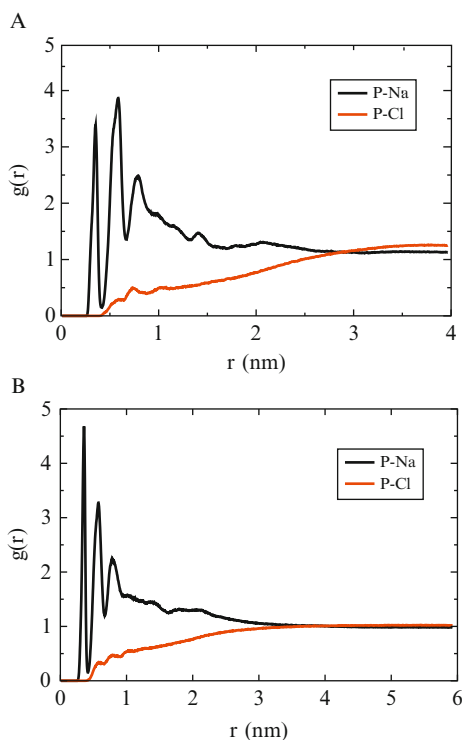


Figure 20.2 Illustration of how finite size artifacts lead to concentration mismatches in the bulk: (A) RNA phosphate-counterion radial distribution functions for the Tar–Tar[★] complex in an 80 Å box of 800 mM NaCl. The anion concentration is $\sim 15\%$ different than the cation concentration at large separations. (B) RNA phosphate-counterion radial distribution functions for the Tar–Tar[★] complex in a 120-Å box of 800 mM NaCl. The anion and cation concentrations agree with each other to within 1% at large radii. Taken from Chen *et al.* (2009).

exclusion of coions in the immediate vicinity of the macroion. The even split between counterion accumulation and coion exclusion observed in these simulations is consistent with equilibrium dialysis experiments of polyelectrolytes at high concentrations ($\sim 1\text{ M}$) of excess monovalent salt (Strauss *et al.*, 1967). In the smaller simulation cell, there is a spurious boundary accumulation of the coions that have been depleted from the vicinity of the RNA. This spurious boundary accumulation arises because the small simulation cell cannot support a proper bulk environment and is unable to absorb the accumulated coions into a bulk-like reservoir.

The anomalous repartitioning described above can also be quantified in terms of preferential interaction coefficients, Γ_+ and Γ_- , respectively. Details of how these coefficients are calculated have been published (Chen *et al.*, 2009). At high concentrations (such as 800 mM), the split between Γ_+ and Γ_- should be roughly even (Strauss *et al.*, 1967). Conversely, for lower concentrations macroscopic thermodynamics requires that $\Gamma_+ > \Gamma_-$ (with the discrepancy increasing with decreasing concentration) and for higher concentrations $\Gamma_+ < \Gamma_-$; that is, excluded volume considerations lead to the dominance of coion depletion over counterion accumulation for salt concentrations that are greater than 1 M . Table 20.1 shows how Γ_+ and Γ_- vary with the size of the simulation cell. Contrary to the expectations from macroscopic thermodynamics, Table 20.1 demonstrates that the smaller box exhibits a smaller degree of coion exclusion and a concomitant higher degree of counterion inclusion, and this is entirely a consequence of the artificially truncated ionic atmosphere of the smaller simulation box.

In the quantification of ion accumulation around a macroion, it is assumed that the macroion is in contact with a thermodynamic bath of bulk electrolyte from which individual ions can be added or subtracted without significantly altering the properties of the bath itself. One way to assay the existence of two independent ion populations within the simulation is to quantify net-charge fluctuations as a function of time within spherical shells around the macroion. Ideally, the shells farthest from the macroion should correspond to the bulk population and should not be strongly coupled to ion fluctuations within the ionic atmosphere of the

Table 20.1 Preferential interaction coefficients calculated as a function of box size for the Tar–Tar* systems and a target concentration of 800 mM of NaCl

Box size (length of each side of the cubic box) (Å)	No. of excess ion pairs	Calculated Γ_+	Calculated Γ_-
80	236	19.6 ± 1.0	-9.3 ± 1.5
120	792	15.0 ± 2.3	-14.8 ± 2.7

macroion. Panels A and B of Fig. 20.3 show the cumulative net charge contained within spherical shells of varying radii from the macromolecule, averaged over 100 ps intervals. This measure is calculated as the average RNA-counterion cumulative distribution function (CDF) minus the RNA-coion CDF. Since the macromolecule bears a net charge of $-30e$,

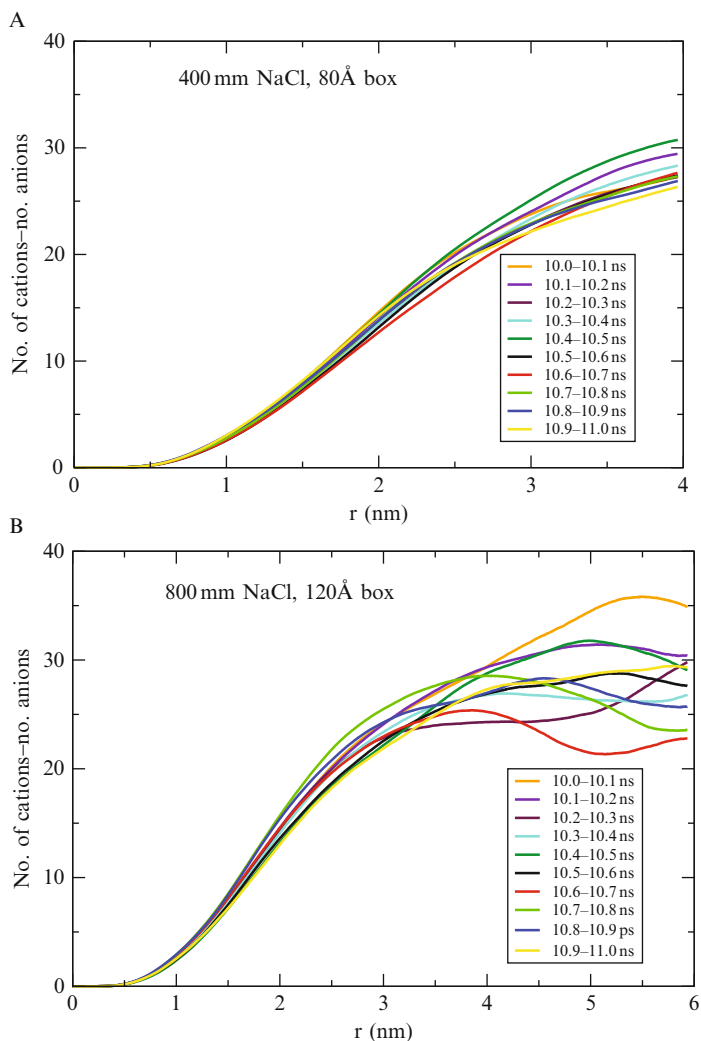


Figure 20.3 Box size dependencies of the accumulated net charge around the Tar-Tar* system. The quantity on the ordinate is plotted against shells of increasing radii and different families of curves are shown for different time intervals within the simulations.

the net-charge CDF for the ions around the RNA should reach $+30e$ at the length scale at which the charge is completely compensated by the ionic atmosphere. It can be seen that in the 80 Å box (panel A of Fig. 20.3), the net-charge CDF only reaches $+30e$ at the very edge of the box, with no visible plateau region. The lack of a plateau in the net-charge CDF is problematic because it indicates that even the solvent fraction farthest from the macroion is still technically part of the “local atmosphere” and not a truly independent population. Furthermore, the charge fluctuations appear to be coupled on all length scales, which is evident by the similarity of the net-charge CDFs both close to and far away from the macroion. In contrast, panel B of Fig. 20.3 shows that the situation is quite different for the 120 Å box. First, a clear plateau region emerges from 40 to 60 Å in the 120 Å box. In addition, this region experiences large net-charge fluctuations that are completely uncoupled from what is occurring very close to the macroion, indicating that there exists a bulk-like population of electrolyte that lies beyond the local ionic atmosphere around the macroion.

Finite size artifacts also impact the kinetics of ion relaxation around the macroion. Converged simulations of ion accumulation around a macroion can only be determined if the simulation is much longer than the intrinsic timescale of ion exchange between the local and bulk partitions. In prior work (Chen and Pappu, 2007b), it was shown that the reorganization time of bulk electrolytes in the absence of a macromolecule was approximately 100 ps. To assess how this timescale was altered by the presence of the RNA, we employed a measure that is similar to that used by Ponomarev *et al.* (2004) and earlier by Cheatham and Young (2000). They quantified convergence rates by comparing the similarity between the atom-resolved distributions of nucleic acid-counterion contacts between the two identical strands of the Drew–Dickerson dodecamer. In the simulations that employ only neutralizing Na^+ counterions (i.e., no excess salt), Ponomarev *et al.* found that the ions exhibited very slow relaxation, characterized by a Pearson correlation coefficient (PCC) of 0.69 at 10 ns and 0.92 at 60 ns, leading to the extrapolation that ~ 100 ns would be required for full convergence (PCC of 0.97) of the ionic atmosphere (Ponomarev *et al.*, 2004).

The Tar–Tar \star sequence as well as the A-form RNA and B-form DNA equivalents of Tar–Tar \star (see Fig. 20.3) do not contain palindromes. However, we have data from multiple independent simulations and the analysis of Ponomarev *et al.* can be repeated using data from pairs of simulations. The equivalent test is to compare the residue-resolved ion-contact distributions between replicate trajectories. Such comparisons are a stringent test if each simulation was separately initialized with randomized ion starting positions and velocities. The PCCs for the A-form RNA simulations are shown in Fig. 20.4. Error bars denote the standard errors across the six possible pairwise comparisons across four independent trajectories, each of

20 ns length. Within a 1 ns interval, the two trajectories have a PCC of 0.67, and this reaches 0.94 by 15 ns, nearly an order of magnitude faster than observed in the Drew–Dickerson dodecamer simulations of Ponomarev *et al.* Therefore, the use of only neutralizing counterions results in significantly retarded ion reorganization kinetics compared to simulations employing large numbers of excess ions. This observation of slower relaxation kinetics suggests that spurious ion–ion correlations are introduced into systems with small numbers of ions. This result is explained as follows: Counterions are predisposed to be in the vicinity of the macroion. If only neutralizing counterions are present, then there is no reservoir with which ions can be exchanged. In the absence of such a reservoir, ions within the local environment around the macroion have to exchange with each other rather than the bulk. This local exchange is inevitably slow because these ions are also restricted to be in the vicinity of the macroion.

The preceding analysis shows that there is a distinct lower bound for the size of the simulation cell. Below this bound, finite size artifacts heavily bias the thermodynamics and kinetics of ion accumulation. Therefore, it seems unadvisable to design simulations using neutralizing counterions alone, especially if the objective is to understand ion-mediated conformational transitions or to understand how ions interact with a macroion. The minimal number of excess ions needed to simulate 800 mM excess monovalent salt was nearly identical (~ 800 pairs) for all of the systems studied (32 nt A-RNA, B-DNA, and the Tar–Tar[★] RNA kissing-loop complex), indicating a simple dependence on net macroion charge. For other macromolecular solutes or other concentrations of excess salt, a box-size

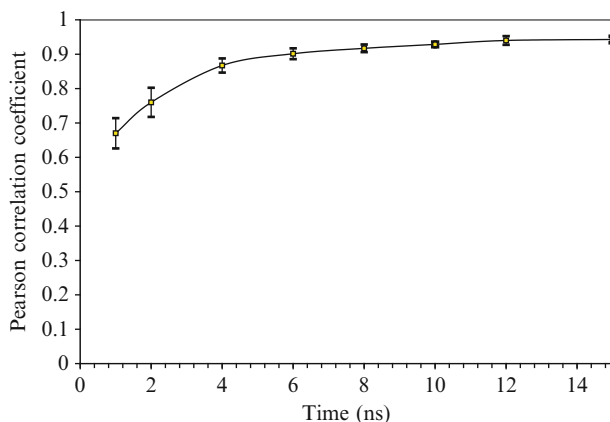


Figure 20.4 Person correlation coefficients quantifying the similarities in ion occupancies around equivalent atoms of the A-form RNA system between different pairs of simulations.

“titration” needs to be performed, with the macroion–ion RDFs providing a quick test to assess whether a reasonably sized ion bath has been used. Finally, a variant of the test suggested by [Ponomarev *et al.* \(2004\)](#), performed using data from independent trajectories with randomized ion starting positions, provides a useful measure of how long a trajectory must be run in order to obtain converged statistics.

3. ON THE USE OF “NEUTRALIZING COUNTERIONS ONLY” IN ALL-ATOM SIMULATIONS

There are several recurring justifications regarding the utility of including only neutralizing counterions in all-atom simulations. The primary justification is one of practicality; that is, to maximize trajectory length by minimizing the volume of bulk solvent included in the central simulation cell. However, the above analysis shows that, in addition to suffering from finite size artifacts, counterion relaxation times are at least an order of magnitude slower when compared to simulations with sufficiently large boxes and clear local–bulk ion partitioning. The second, less-publicized justification is in response to the spurious salt cluster and lattice formation which has been observed to occur in simulations utilizing the AMBER force-field at concentrations well below the solubility limit ([Auffinger *et al.*, 2007](#); [Chen and Pappu, 2007a](#)). In these cases, the omission of the coion is used to mask cluster and lattice formation. However, these artifacts have since been shown to be the result of incorrect ion parameter adaptations, with two independently proposed fixes that eliminate this problem ([Chen and Pappu, 2007a](#); [Joung and Cheatham, 2008](#)). Lastly, there is an analytical solution to the Poisson–Boltzmann equation for the polyelectrolyte cell model, which is only possible for the case of no excess salt. Simulations including only neutralizing counterions are therefore only appropriate in cases where a direct comparison to such theories is desirable.

4. RESULTS FROM SIMULATIONS OF CANONICAL A-FORM RNA AND B-FORM DNA HELICES

Recent simulation work ([Chen *et al.*, 2009](#)) focused on comparative quantification of the accumulation of different types of group I monovalent cations around the Tar–Tar[★] kissing loop. The analysis identified three salient features regarding the distributions of accumulated ions around Tar–Tar[★]: (1) The ionic atmosphere can be quantified in terms of free energy layers around the RNA using the methods described in detail below. The most favorable free energy levels ($\Delta G \leq -1.5$ kcal/mol) are occupied

by partially dehydrated ions. The first hydration shell statistics of these ions are markedly different from those in the bulk and for smaller cations, the extent of dehydration is greater because favorable water–ion interactions are replaced by favorable interactions with the Tar–Tar⁺ complex. The guanine tract in the Tar hairpin facilitates this partial dehydration. (2) There is asymmetry in the extent of cation accumulation around each hairpin. However, the degree of the observed asymmetry is smallest for Na⁺ and largest for Cs⁺ because the smaller cations can form a belt of positive-charge density across the loop–loop interface. (3) Finally, in the bulk-like free energy intervals ($\Delta G > -0.5$ kcal/mol), there remain differences in cation accumulation that are explained in terms of the generic differences between the corresponding 1:1 electrolytes (NaCl, KCl, and CsCl). In particular, at high concentrations (ca. 800 mM), the degree of nonideality varies with cation size and is more prominent for the smaller cations. These collective observations were used to rationalize experimental measurements that demonstrate cation specificity for the free energy of complex formation (Lambert *et al.*, 2009). Recent work (Semichaevisky *et al.*, 2009) showed that asymmetry in counterion accumulation across the loop–loop interface is a characteristic that is shared among a wide variety of kissing-loop complexes. They hypothesized that this asymmetry in counterion accumulation facilitates rapid dissociation kinetics required in the function of naturally occurring antisense RNA feedback loops. This hypothesis needs further testing because of the cation specificity observed in ion asymmetric ion accumulation.

The results summarized above were obtained through the analysis of spatial free energy density profiles using the following methods. Spatial distribution functions for ion occupancies around the macroion are computed on a cubic grid. These distribution functions can be converted to spatial free energy density profiles using the relation:

$$\Delta G(x, y, z) = -RT \ln \left[\frac{\rho(x, y, z)}{\rho_0} \right] \quad (20.1)$$

here, $\rho(x, y, z)$ is the density within each 1 Å³ cell, ρ_0 is the bulk ion density as measured in the periphery of the simulation cell, $R = 1.987 \times 10^{-3}$ kcal/mol-K, and T is the simulation temperature. The resulting free energy contours provide an assessment of the spatial variations of the strengths of ion interactions with the macroion. The cumulative ion numbers within each contour cutoff are calculated by summation of $\rho(x, y, z)$ over all the cells that result in a ΔG value above a specified value. The relationship shown in Eq. (20.1) is based on the fact that, in an equilibrium simulation, the distributions of ions in local and bulk partitions obey Boltzmann statistics. The ratio within the square brackets of Eq. (20.1) quantifies the relative probability—vis-à-vis bulk—of realizing a specific ion density at a

particular location around the RNA. If the ratio is unity, then the logarithm of the ratio is zero, thus setting the bulk milieu as the reference state. Deviations from unity are either due to ion depletion or accumulation and the quantity shown in Eq. (20.1) quantifies the free energy change associated with depleting or accumulating ions vis-à-vis the bulk.

We wished to know if the features of ionic atmosphere around the Tar–Tar[★] were generalizable to typical nucleic acid architectures such as A- or B-form helices. We performed simulations of both an A-form RNA helix and a B-form DNA helix, each 32 nt/16 bp long, and containing nearly identical sequences as found in the Tar–Tar[★] kissing-loop complex (Fig. 20.1). The only modification to the sequences is that the loop region has been altered to form canonical duplexes instead of two separate hairpins; for the B-DNA duplex, the uracil residues have been substituted with thymine residues and ribose rings of RNA were replaced with deoxyribose rings. These systems were simulated under identical conditions as used in the Tar–Tar[★] kissing-loop study, that is, 800 mM excess NaCl and KCl, with 792 excess ion pairs in 120 Å boxes. The nucleic acids were modeled using the AMBER-99 forcefield (Cornell *et al.*, 1995) employing the GROMACS implementation of Sorin and Pande (2005). The rigid three-site TIP3P model (Jorgensen *et al.*, 1983) was used to simulate water molecules. Ions were modeled using the parameters of Åqvist (1990) according to the approach proposed by Chen and Pappu (2007a).

Figure 20.5 shows the spatial free energy density profiles for the A- and B-form systems, respectively. Counterion densities were visualized by creating free-energy isocontours at -0.5 , -1.0 , -1.5 , and -2.0 kcal/mol intervals, identical to the procedure described in previous work (Chen *et al.*, 2009). The integrated numbers for the cumulative number of counterions accumulated within a free energy interval is calculated as a function of free energy cutoff and plotted in Fig. 20.6.

It is known that the change in the melting temperature resulting from a change in the monovalent salt concentration is higher for A-RNA duplexes than B-DNA duplexes (Nakano *et al.*, 1999), whereas the salt dependence of the Tar–Tar[★] complex is even higher still (Lambert *et al.*, 2009). Therefore, we expected a systematic trend whereby Tar–Tar[★] accumulates more counterions than A-form RNA, which in turn accumulates more counterions than B-DNA—across all free energy cutoffs. Additionally, oligomers of B-DNA are known to exhibit identical stabilities in equimolar buffers of NaCl and KCl from 0.5 to 600 mM excess salt (Nakano *et al.*, 1999). Since counterion specificity in Tar–Tar[★] is found to result from differences in the number of highly favorable ($\Delta G < -2.0$ kcal/mol) accumulated counterions, it is expected that the B-form DNA will exhibit minimal differences between Na⁺ and K⁺ in this free energy regime.

We find that both of the canonical helices exhibit less accumulation within all free energy contours when compared to the RNA kissing-loop

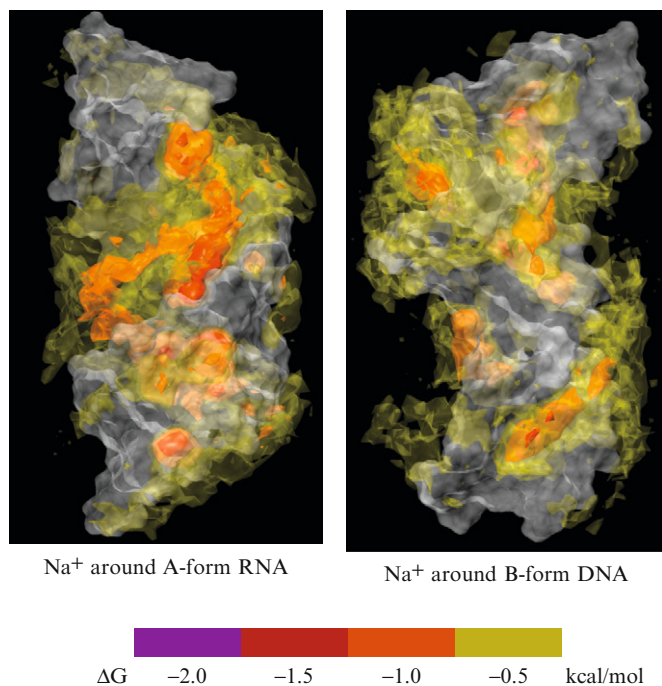


Figure 20.5 Spatial free energy density profiles. Spatial free energy density profiles for counterion accumulation around the A-form RNA (*left*) and B-form DNA (*right*). Free-energy isocontours (surfaces of zero thickness at specific free-energy values) are colored according to the key shown at the bottom of the figure. The volume encompassed by each isocontour contains regions of more negative free energy density than prescribed by the isocontour itself.

system, in the order $\text{Tar-Tar}^{\star} > \text{A-RNA} > \text{B-DNA}$ (Fig. 20.6). Although all three structures systematically accumulate more Na^+ counterions than K^+ counterions at any free energy cutoff, for B-DNA there is negligible counterion accumulation in the most favorable range ($\Delta G < -2.0$ kcal/mol), which is the free energy interval that we expect to be most important for specificity in counterion accumulation. Differences in counterion accumulation that go beyond the lowest and intermediate free energy contours ($\Delta G > -1.0$ kcal/mol) reflect the nonidealities in the bulk electrolytes at high concentrations. The region of highest accumulated counterion density in the A-form RNA lies within the major groove, along the cylindrical axis of the helix. In contrast, the B-form DNA exhibits more diffuse layers of counterion accumulation, which are mostly situated along the major and minor grooves. These conclusions are drawn based on the comparative spatial free energy density profiles shown in Fig. 20.5 where panel A shows the profile for Na^+

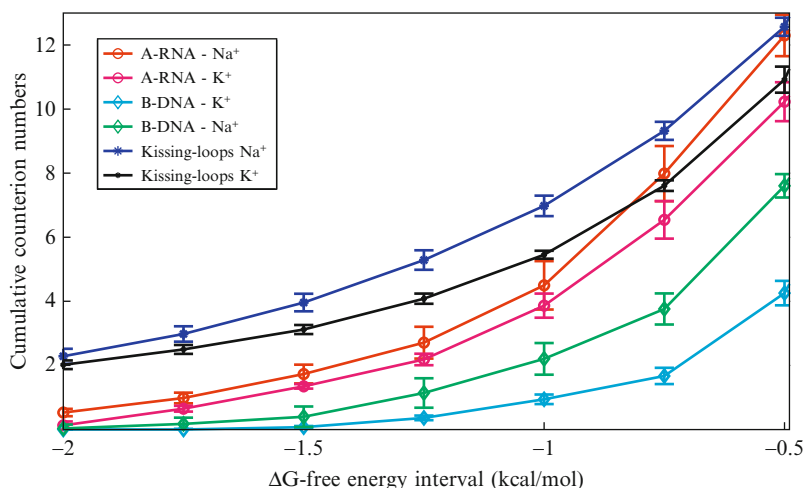


Figure 20.6 Comparative cumulative distribution functions to assess the amount of counterion accumulation around the Tar–Tar[★] complex, A-form RNA, and B-form DNA, respectively. Data were obtained from molecular dynamics simulations with explicit representations of water molecules and ions. The ordinate quantifies the number of counterions accumulated around the macroions within contours that have ΔG less than or equal to the value on the abscissa.

accumulation around the A-form RNA and panel B the corresponding profile for accumulation around B-form DNA. This reflects the inherent difference in helix architecture, as the A-form helix possesses an accessible “open core” compared to the completely occluded core of the canonical B-form helix. Both A-RNA and B-DNA exhibit prominent “spines” of moderate to weakly favorable ($-1.0 \leq \Delta G \leq -0.5$ kcal/mol) counterion accumulation threaded along the helical grooves. This is split roughly equally across major and minor grooves in B-DNA but is predominantly in the major groove in A-RNA.

The tracking of counterion density with nucleotide sequence is much more pronounced in A-form RNA than in the B-form DNA. The region of most favorable counterion accumulation lies directly adjacent to the tract of tandem purines, including the three tandem guanines that were found to be key in the accumulation of ions at the loop–loop interface of the Tar–Tar[★] complex. This favorable counterion accumulation lies along the cylindrical axis of the A-form RNA and ends abruptly at the center of the structure, which is where the purine tract ends. It is at this point that the major groove develops a “spine” of moderate to weakly favorable ($-1.0 \leq \Delta G \leq -0.5$ kcal/mol) accumulated counterions, reminiscent of the diffuse “belt” found straddling the loop–loop interface in our previous characterization of the Tar–Tar[★] kissing-loop system. The similarity

between the kissing-loop complex and the A-RNA duplex is also evident in the quantification of number of accumulated counterions, in that the kissing-loop complex and A-RNA all converge to roughly the same number at the weakest free energy cutoff examined ($\Delta G \leq -0.5$ kcal/mol) whereas the B-DNA does not. Therefore, it would appear that the architecture of the A-form RNA is inherently predisposed for particular patterns of counterion accumulation, which are amplified in the context of a kissing-loop complex.

In recent work, Savelyev and Papoian examined the relative affinities of 15 Na^+ and 15 K^+ ions to a 16 bp DNA helix, in which a systematic preference for accumulation of Na^+ over K^+ is also observed (Savelyev and Papoian, 2008). The authors concluded that the accumulation of Na^+ is favored due to the trade-off between the more favorable electrostatic potential accessible to smaller ions, which outweighs the greater desolvation cost for partial dehydration. Our results for the Tar-Tar \star system and the two systems studied here agree with the trends identified by Savelyev and Papoian. However, the proposed mechanisms for cation specificity are different: whereas the ionic potential used in our studies result in K-Cl pairing constants that are systematically less favorable than Na-Cl pairing constants (Chen and Pappu, 2007b), the interionic potential employed by Savelyev and Papoian predict the reverse situation, resulting in a coion competition effect which sequesters K^+ ions from the vicinity of the macroion. These discrepancies in interpretation call for systematic comparative studies, and these are ongoing.

5. COMPARISON WITH PREDICTIONS OF THE NONLINEAR POISSON-BOLTZMANN EQUATION

Atomistic simulations based on explicit solvent models are informative and yield important insights regarding the origins for specificity in counterion accumulation. Spatial free energy density profiles are useful devices for quantifying the intricate details of the liquid-like organization of ions around the macroion. Integrals of these profiles yield numeric values for the numbers of ions in different layers, and these quantities are useful for extracting a range of colligative properties. However, as discussed in the section on simulation setups, it is clear that large simulation cells are needed in order to obtain thermodynamically accurate quantification of the ionic atmospheres around RNA and DNA molecules. To be truly effective, we need to be able to extend such calculations to a range of RNA systems, many of which are considerably larger than 32 nucleotides. Furthermore, to generate insights regarding driving forces for conformational changes and ligand binding, it is important to be able to simulate the salt dependence of

ion binding and the coupling between ion-interactions and conformational fluctuations. The ability to simulate such systems at a variety of salt concentrations would help elucidate the nature of the strong coupling between ion-macromolecule interactions and the macromolecular conformational equilibria. It is difficult to address these issues through simplistic scaling up of simulations based on explicit representations of solvent molecules. Continuum solvent models possess an intrinsic advantage because the effects of solvation can be captured implicitly, thereby allowing for the savings in computational expense. However, the issue of accuracy vis-à-vis explicit, all-atom simulations remains unresolved. Accordingly, we assess the utility of the most commonly used continuum solvation model by comparing spatial free energy density profiles and their integrals to those obtained from simulations with explicit solvent for the Tar-Tar*, A-form RNA, and B-form DNA systems discussed above.

Analysis of ion atmospheres around highly charged macromolecules has traditionally been performed using numerical solutions to the nonlinear Poisson-Boltzmann (P-B) equation (Anderson and Record, 1980; Bai *et al.*, 2007; Baker, 2004), in which the macromolecule is approximated as a collection of point charges embedded in a low dielectric cavity surrounded by a high-dielectric solvent. This approach utilizes the precise three-dimensional structure of the macromolecule (albeit in a static sense). We would not expect such a framework to capture subtleties, which are dependent on the partial dehydration of ions.

The spatial free energy density plots and the numbers of accumulated counterions as a function of free energy cutoff obtained using the APBS implementation (Baker *et al.*, 2001; Dolinsky *et al.*, 2004) of the nonlinear Poisson-Boltzmann equation are presented in Figs. 20.7 and 20.8, respectively. Calculations utilized a bulk salt concentration of 800 mM salt, with the hydrated radii of Na⁺, K⁺, and Cl⁻ set to 3.33, 4.93, and 4.42 Å, respectively. The solute and solvent dielectric constants were set to 20 and 78.54, respectively. The solute-solvent dielectric interface was calculated using a molecular surface definition using a solvent probe radius of 1.4 Å. The nonlinear Poisson-Boltzmann equation was solved on a grid of 257 × 257 × 257 grid points, corresponding to a grid spacing of 0.469 Å. The most representative single structure for each system was extracted from the molecular-dynamics trajectories through cluster analysis using the GROMACS *g_cluster* utility, which identified the structural snapshot most equidistant from all other snapshots in terms of root mean square deviation (RMSD). At first glance, the numbers of accumulated counterions in Fig. 20.8 are rather similar to the values from the all-atom simulations (Fig. 20.6). Both the all-atom molecular dynamics and the P-B calculations agree that B-DNA accumulates substantially fewer highly favorable ($\Delta G \leq -1.5$ kcal/mol) counterions than A-RNA. However, the P-B calculations are unable to capture the greater propensity for

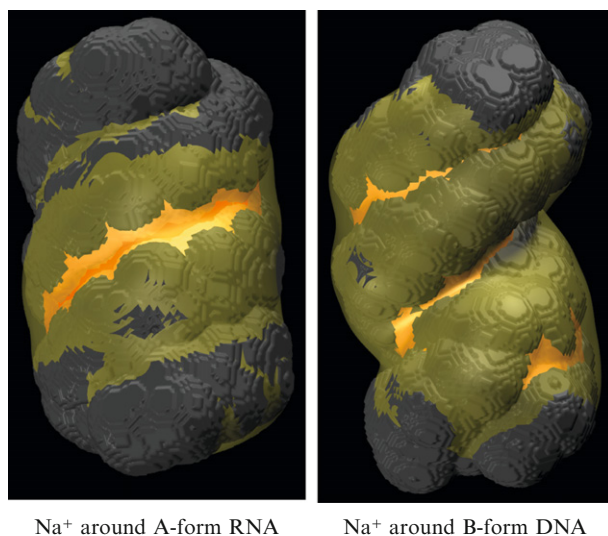


Figure 20.7 Spatial free energy density profiles obtained using data from P-B calculations. Grid sizes of 1 \AA^3 were used in these calculations. The color-coding for the free energy levels is based on the key shown at the bottom of Fig. 20.5.

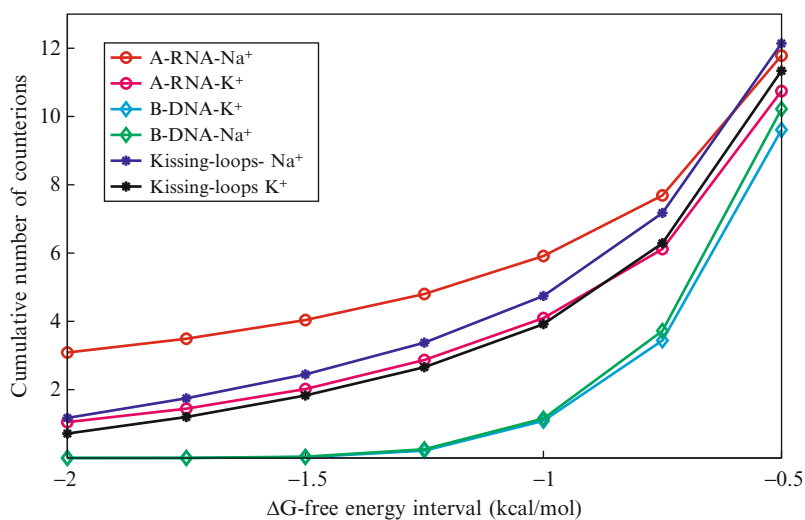


Figure 20.8 Comparative cumulative distribution functions to assess the amount of counterion accumulation around the Tar-Tar^{*} complex, A-form RNA, and B-form DNA, respectively. Data were obtained from P-B calculations and are to be compared with those shown in Fig. 20.6.

counterion accumulation around the Tar–Tar[★] kissing loop as compared to the A-form RNA.

While the trends for the integrals of spatial density profiles are roughly similar to each other (cf. Figs. 20.6 and 20.8), close examination of the spatial free energy density profiles reveals that the trends in counterion accumulation as predicted by the P–B calculations do not arise from the same regions as in the molecular dynamics simulations. Specifically, the interior of both the Tar–Tar[★] kissing loops as well as that of the A-form RNA are completely occluded by the Stern layer, which was set to the hydrated radii of the ions. For all three structures, only a narrow band within the helical grooves is accessible to the ions in the P–B calculations. For B-form DNA, both the major and minor grooves are accessible to weak to moderately accumulated counterions ($-1.0 \leq \Delta G \leq -0.5$ kcal/mol) while only the deeper major groove of the A-form RNA is accessible to the favorable counterions at -1.5 kcal/mol. The greater accumulation around the A-form RNA in the P–B calculation is therefore a simple consequence of the major groove being deeper than that of B-DNA, which agrees with X-ray scattering measurements and P–B calculations (Pabit *et al.*, 2009). Similarly, the accumulation differences between K⁺ versus Na⁺ are captured solely on account of the larger Stern layer for K⁺, resulting in shallower grooves. It is notable that this aspect of counterion accumulation, although incomplete, is still in agreement with the “spines” and “belts” observed in the molecular dynamics simulations, which we have previously determined to consist of fully hydrated ions.

For the Tar–Tar[★] kissing loops, the P–B calculations are unable to discern their propensity to accumulate counterions accumulation at the loop–loop interface (data not shown). This is because the fully hydrated ions as defined by the Stern layer cannot penetrate into the central cation “binding pocket” (data not shown). Similarly, the axial spine of counterion density observed in the A-RNA helix (Fig. 20.5) is not captured by the P–B calculation (Fig. 20.7). No noticeable sequence specificity is observed in the counterion accumulation patterns in the P–B calculations, even though the sequence effects are explicitly represented in the P–B calculation through the appropriate geometry and assignment of point-charges. This is because the sequence specificity observed in the molecular dynamics simulations usually involves first shell interactions of base moieties with partially dehydrated ions, which cannot be accurately represented in the P–B framework.

Defining a Stern layer using hydrated radii of the ions is justified in the P–B framework, since ions must be fully hydrated in order for the uniform dielectric approximation to be valid. If ions were allowed to penetrate this layer, then additional physics would have to be introduced to account for ion–ion correlations, partial dehydration, and other considerations that fall well outside the realm of applicability for standard continuum solvent models. To generalize the P–B framework, new approaches are needed

and promising results have been reported along these lines (Xu *et al.*, 2005; Yi *et al.*, 2008). These include the size-modified Poisson–Boltzmann theory (Borukhov *et al.*, 1997), in which a lattice-gas partition function is introduced to limit the maximum possible ionic density (Chu *et al.*, 2007), as well as the tightly bound ion theory (Tan and Chen, 2005), in which additional ion–ion correlations are introduced in areas of high electrostatic potential.

Spatial free energy density profiles extracted from analysis of data obtained using simulations based on explicit representations of solvent molecules suggest that hydrated ions make up a majority of the atmosphere around RNA and DNA structures (Fig. 20.5). As a result, integrals of spatial free energy density profiles obtained from P–B calculations yield numerical values that are in the right range and show many—although not all—of the trends obtained from integrals of profiles from more expensive simulations. This bodes well for the use of modified P–B models that have the ability of incorporating ion–ion correlations or the effects of partially hydrated ions. These results also suggest that the use of grand canonical ensemble Monte Carlo (GCMC) simulations (Vitalis *et al.*, 2004) combined with empirical implicit solvation models such as the recently developed ABSINTH model (Vitalis and Pappu, 2009) might provide a way to combine the virtues of solvent coarse graining and explicit inclusion of ion–ion correlations and many, if not all the necessary effects of partially desolvated ions. In fact, initial results applying a slightly modified ABSINTH model have yielded promising results for the Tar–Tar[★] system (Chen and Pappu, unpublished data) and more work is necessary in this direction.

ACKNOWLEDGMENT

This work was supported by grant MCB-0718924 from the National Science Foundation (RVP).

REFERENCES

- Anderson, C. F., and Record, M. T. Jr. (1980). The relationship between the Poisson–Boltzmann model and the condensation hypothesis: An analysis based on the low salt form of the Donnan coefficient. *Biophys. Chem.* **11**, 353–360.
- Anderson, C. F., and Record, M. T. Jr. (1995). Salt–nucleic acid interactions. *Annu. Rev. Phys. Chem.* **46**, 657–700.
- Åqvist, J. (1990). Ion water interaction potentials derived from free-energy perturbation simulations. *J. Phys. Chem.* **94**, 8021–8024.
- Auffinger, P., Cheatham, T. E. III, and Vaiana, A. C. (2007). Spontaneous formation of KCl aggregates in biomolecular simulations: A force field issue? *J. Chem. Theory Comput.* **3**, 1851–1859.

- Bai, Y., Greenfield, M., Travers, K. J., Chu, V. B., Lipfert, J., Doniach, S., and Herschlag, D. (2007). Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.* **129**, 14981–14988.
- Baker, N. A. (2004). Poisson-Boltzmann methods for biomolecular electrostatics. *Numer. Comput. Methods* **383**(Pt D), 94–118.
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041.
- Borukhov, I., Andelman, D., and Orland, H. (1997). Steric effects in electrolytes: A modified Poisson-Boltzmann equation. *Phys. Rev. Lett.* **79**, 435–438.
- Chang, K.-Y., and Tinoco, I. Jr. (1997). The structure of an RNA “kissing” hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.* **269**, 52–66.
- Cheatham, T. E. III, and Young, M. A. (2000). Molecular dynamics simulations of nucleic acids: Successes, limitations and promise. *Biopolymers* **56**, 232–256.
- Chen, A. A., and Pappu, R. V. (2007a). Parameters of monovalent ions in the AMBER-99 forcefield: Assessment of inaccuracies and proposed improvements. *J. Phys. Chem. B* **111**, 11884–11887.
- Chen, A. A., and Pappu, R. V. (2007b). Quantitative characterization of ion pairing and cluster formation in strong 1:1 electrolytes. *J. Phys. Chem. B* **111**, 6469–6478.
- Chen, A. A., Draper, D. E., and Pappu, R. V. (2009). Molecular simulation studies of monovalent counterion-mediated interactions in a model RNA kissing loop. *J. Mol. Biol.* **390**, 805–819.
- Chu, V. B., Bai, Y., Lipfert, J., Herschlag, D., and Doniach, S. (2007). Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory. *Biophys. J.* **93**, 3202–3209.
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
- Das, R., Kwok, L. W., Millett, I. S., Bai, Y., Mills, T. T., Jacob, J., Maskel, G. S., Seifert, S., Mochrie, S. G. J., Thiyagarajan, P., Doniach, S., Pollack, L., and Herschlag, D. (2003). The fastest global events in RNA folding: Electrostatic relaxation and tertiary collapse of the tetrahymena ribozyme. *J. Mol. Biol.* **332**, 311–319.
- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667.
- Draper, D. E. (2004). A guide to ions and RNA structure. *RNA* **10**, 335–343.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Phys. Chem.* **79**, 926–935.
- Joung, I. S., and Cheatham, T. E. III (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041.
- Laing, L. G., Gluick, T. C., and Draper, D. E. (1994). Stabilization of RNA structure by Mg ions specific and non-specific effects. *J. Mol. Biol.* **237**, 577–587.
- Lambert, D., Leipply, D., Shiman, R., and Draper, D. E. (2009). The influence of monovalent cation size on the stability of RNA tertiary structures. *J. Mol. Biol.* **390**, 791–804.
- Nakano, S.-I., Fujimoto, M., Hari, H., and Sugimoto, N. (1999). Nucleic acid duplex stability: Influence of base composition on cation effects. *Nucleic Acids Res.* **27**, 2957–2965.

- Pabit, S. A., Qiu, X., Lamb, J. S., Li, L., Meisburger, S. P., and Pollack, L. (2009). Both helix topology and counterion distribution contribute to the more effective charge screening in dsRNA compared with dsDNA. *Nucleic Acids Res.* **37**, 3887–3896.
- Pan, J., Thirumalai, D., and Woodson, S. A. (1999). Magnesium-dependent folding of self-splicing RNA: Exploring the link between cooperativity, thermodynamics, and kinetics. *Proc. Natl. Acad. Sci. USA* **96**, 6149–6154.
- Perez-Salas, U. A., Rangan, P., Krueger, S., Briber, R. M., Thirumalai, D., and Woodson, S. A. (2004). Compaction of a bacterial group I ribozyme coincides with the assembly of core helices. *Biochemistry* **43**, 1746–1753.
- Ponomarev, S. Y., Thayer, K. M., and Beveridge, D. L. (2004). Ion motions in molecular dynamics simulations on DNA. *Proc. Natl. Acad. Sci. USA* **101**, 14771–14775.
- Rangan, P., and Woodson, S. A. (2003). Structural requirement for Mg^{2+} binding in the group I intron core. *J. Mol. Biol.* **329**, 229–238.
- Savelyev, A., and Papoian, G. A. (2008). Polyionic charge density plays a key role in differential recognition of mobile ions by biopolymers. *J. Phys. Chem. B* **112**, 9135–9145.
- Semichaevsky, A. V., Marlowe, A. E., and Yingling, Y. G. (2009). Comparing ion distributions around RNA and DNA helical and loop-loop motifs. *Mater. Res. Soc. Symp. Proc.* **1130**, W05.
- Smolin, N., and Winter, R. (2004). Molecular dynamics simulations of staphylococcal nuclease: Properties of water at the protein surface. *J. Phys. Chem. B* **108**, 15928–15937.
- Sorin, E. J., and Pande, V. S. (2005). Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* **88**, 2472–2493.
- Strauss, U. P., Helfgott, C., and Pink, H. (1967). Interactions of polyelectrolytes with simple electrolytes. II. Donnan equilibria obtained with DNA in solutions of 1–1 electrolytes. *J. Phys. Chem.* **71**, 2550–2556.
- Takamoto, K., Das, R., He, Q., Doniach, S., Brenowitz, M., Herschlag, D., and Chance, M. R. (2004). Principles of RNA compaction: Insights from the equilibrium folding pathway of the P4–P6 RNA domain in monovalent cations. *J. Mol. Biol.* **343**, 1195–1206.
- Tan, Z.-J., and Chen, S.-J. (2005). Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *J. Chem. Phys.* **122**, 044903.
- Vitalis, A., and Pappu, R. V. (2009). ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *J. Comput. Chem.* **30**, 673–699.
- Vitalis, A., Baker, N. A., and McCammon, J. A. (2004). ISIM: A program for grand canonical Monte Carlo simulations of the ionic environment of biomolecules. *Mol. Simul.* **30**, 45–61.
- Woodson, S. A. (2008). RNA folding and ribosome assembly. *Curr. Opin. Chem. Biol.* **12**, 667–673.
- Xu, D. R., Greenbaum, N. L., and Fenley, M. O. (2005). Recognition of the spliceosomal branch site RNA helix on the basis of surface and electrostatic features. *Nucleic Acids Res.* **33**, 1154–1161.
- Yi, M., Nymeyer, H., and Zhou, H. X. (2008). Test of the Gouy-Chapman theory for a charged lipid membrane against explicit-solvent molecular dynamics simulations. *Phys. Rev. Lett.* **101**, 038103–1–4.