THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

# Aqueous Guanidinium–Carbonate Interactions by Molecular Dynamics and Neutron Scattering: Relevance to Ion–Protein Interactions

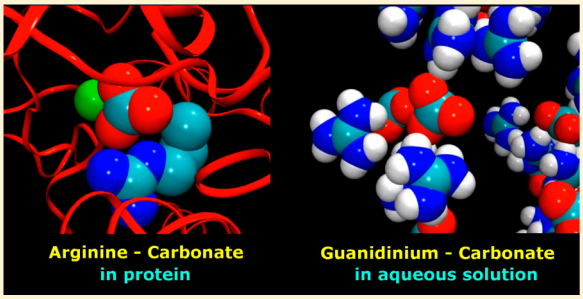
Mario Vazdar,[†,‡] Pavel Jungwirth,[†] and Philip E. Mason*[,†]

[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 16610 Prague 6, Czech Republic

[‡]Division of Organic Chemistry and Biochemistry, Rudjer Bošković Institute, P.O.B. 180, HR-10002 Zagreb, Croatia

Supporting Information

**ABSTRACT:** Guanidinium carbonate was used in this study as a simple proxy for the biologically relevant arginine–carbonate interactions in water. Molecular dynamics (MD) simulations of guanidinium carbonate were performed with nonpolarizible water using two implementations of the ion force fields. In the first, the ions had full charges, while in the second, the ions had reduced charges in order to effectively account for electronic polarization effects of water. The results from the simulations were then compared to data from previous neutron scattering experiments. It was found that there were significant discrepancies between the full charge force field MD simulations and the experimental results due to excessive ion pairing and clustering in the former. In contrast, reducing the ionic charges yields a more regular solution with a simulated structure, which fits well the experimental data.




## INTRODUCTION

Arginine is one of the most fascinating amino acids. With its side chain essentially equating to a propyl group terminated with a cationic guanidinium ion, it is the third largest (by mass) of all of the natural amino acids (second only to tyrosine and tryptophan). The peculiar behaviors of arginine and the isolated guanidium ion have generated a lot of interest in the literature.[1,2] Notably, guanidinium ions in aqueous solution are found to associate with each other, despite both ions bearing a full positive charge.[4−9] It has been suggested that this homoion pairing of the arginine side chains is actually a significant factor both in protein folding and in protein aggregation.[10,11] Further, its interactions with anionic species are thought to be similarly important. For example, arginine is known to make salt bridges with the side chains of glutamate and aspartate.[3] Arginine also interacts directly with small oxyanions such as the dianionic carbonate ion. It is typically found double hydrogen bonded to the carbonate ion as in the example of the highly conserved binding site of many iron transport proteins.[4] Of these, by far the largest and most important family is that of the lactoferrins,[5] where the direct interaction of the arginine with the carbonate ion (chelated to an iron ion) is thought to play a key role in regulating the function of the protein.[6]

A previous MD study[7] suggested that in solutions of guanidinium carbonate, the $Gdm^{+}$–$CO_3^{2-}$ double hydrogen bond type of interaction of the two ions was relatively strong. Indeed, in this study, it was found to be strong enough to lead to nanosecond time scale ion pairs and resulted in aggregated ion structures on the nanometers length scale.[7] The comparison of the molecular dynamics predictions to the neutron scattering data seemed to provide some support for this conclusion.[7]

Here, we re-examine this study in light of new work that suggests that in nonpolarizable simulations containing fully charged ions in a solution with simple point charge water, the ion–ion interactions are significantly overemphasized.[8] Calculations have suggested that in order to more accurately represent ions in simple point charge water simulations, the charge on the ions needs to be reduced by 25% (vide infra) to account for the electronic part of the polarizability of the aqueous solution. A similar "charge scaling" is commonly used in nonpolarizabile MD simulations of ionic liquids.[9] This reduced charge model is called the electronic continuum correction or ECC model throughout this study. It should be noted that reducing the ion charges by 25% is a significant alteration to the force field. Such large changes would be expected to significantly reduce the free energy of hydration of the ECC ions if they were calculated solely on the basis that the system consisting of fixed charges. However, the current ECC approach[10] recognized that there is an additional component to the hydration free energy due to the electronic polarization.

**Received:** October 30, 2012
**Revised:** December 11, 2012
**Published:** December 17, 2012

   dx.doi.org/10.1021/jp310719g | J. Phys. Chem. B 2013, 117, 1844–1848

When calculating the free energies of hydration by this approach, it is therefore necessary to add the electronic Born solvation term (employing the electronic part of the dielectric constant of $\varepsilon_{el} = 1.78$).[10]

It would be of considerable value to determine if such factors are relevant in accurately modeling ion−protein interactions. One way to do this is via the comparison of predictions from MD simulations with direct structural experimental measurements. In this study, guanidinium carbonate was chosen as a test case to examine this point of contest. As mentioned above, guanidinium is a proxy for the side chain of the amino acid arginine, and carbonate is an important biological ligand. This study examines ion pairing in aqueous solutions of guanidinium carbonate via MD simulations with either full charges on ions or scaled charges within the ECC model and then compares the results to previous neutron scattering measurements on the same system.[7] We show that the ECC correction, which leads to a significant reduction in ion pairing and clustering, is crucial for obtaining a quantitative agreement with experiment.

## SIMULATIONS DETAILS

Two independent 50 ns MD simulations (after 1 ns of equilibration) were run from identical starting coordinates. In order for the simulations to be of an identical concentration to the experimental measurements, both simulations contained 48 $Gdm^+$ ions,[11] 24 $CO_3^{2-}$ ions,[7] and 889 water molecules (SPC/E).[12] In the first simulation, the ions bear full formal charge, that is, the charge is +1 for the $Gdm^+$ ion (atom charges: C, 0.64; N, −0.80; H, 0.46) and −2 for the $CO_3^{2-}$ ion (atom charges: C, 0.676; O, −0.892), respectively, as previously used.[7] In the second simulation, electronic polarization effects were accounted for in an effective way by introducing ECC, which is practically realized by rescaling all ionic charges by $1/(\varepsilon_{el})^{1/2}$, where $\varepsilon_{el} = 1.78$ is the electronic part of the static relative permittivity of water (i.e., the scaling factor equals 0.75).[8,10,13] In this way, the $Gdm^+$ ion has a charge of 0.75 (atom charges: C, 0.48; N, −0.60; H, 0.345), and the charge is −1.5 for the $CO_3^{2-}$ ion (atom charges: C, 0.507; O, −0.669).

Periodic boundary conditions were used with long-range electrostatic interactions beyond the nonbonded cutoff of 10 Å, which accounted for using the particle-mesh Ewald procedure.[14] The Berendsen thermostat with a temperature of 300 K was used,[15] and the box length was rescaled to 31.7840 Å; this yielded the correct physical number density (0.101 atoms $Å^{-3}$). The SHAKE algorithm[16] was employed to constrain all bonds containing hydrogen atoms. The time step used in all simulation was set to 1 fs. Molecular dynamics simulations were performed with AMBER 11.[17]

These simulations were then used to calculate the function measured in the previous first-order $^{nat}N/^{15}N$ neutron diffraction with the isotopic substitution (NDIS) experiment.[7] Specifically, the reciprocal space function is

$$ {}^{n}\Delta_N(Q)_{null} = \frac{4.46S_{NO}(Q) + 0.91S_{NN}(Q) + 0.38S_{NC}(Q)}{4.46 + 0.91 + 0.38} - 1 \quad (1) $$

and the real space version of this function is

$$ {}^{n}G_N(r)_{null} = \frac{4.46g_{NO}(r) + 0.91g_{NN}(r) + 0.38g_{NC}(r)}{4.46 + 0.91 + 0.38} - 1 \quad (2) $$

where the subscript *null* represents that these measurements were made on solutions in which the average scattering length of hydrogen was zero. This effectively renders the exchangeable hydrogens "neutron invisible" in these solutions. The superscript *n* represents that this measurement was normalized by dividing through by the sum of the prefactors. There are two chemical types of carbon in these solutions (carbon in carbonate and carbon in guanidinium) and two types of oxygen (oxygen in water and oxygen in carbonate). While these functions can be further resolved, in the current case, it is only useful to do so with the oxygens, of which 7.5% are on the carbonate ($O_{carb}$) and 92.5% are on the water ($O_w$).

The level of solute−solute interaction depends on the competition of the ion−ion, ion−water, and water−water interactions, and hence, the force field used.[18] Some caution over the sensitivity of the observed behavior to the force field used must therefore be exercised. However, we have previously found that in the case of relatively strong interactions (like the $Gdm^+$−$CO_3^{2-}$ interactions in this study), there is a limited sensitivity to the water model chosen.[19,20]

## RESULTS FROM MOLECULAR DYNAMICS

Representative snapshots from the full charges and ECC simulation are shown in Figure 1. From these trajectories, density maps were calculated for the distribution of carbonate around guanidinium (Figure 2). The radial distribution functions of the relevant atoms around the nitrogen atoms in these systems were also calculated. These were then weighted by the neutron scattering prefactors shown in eq 2 (Figure 3) and summed to emulate the experimental neutron scattering result that would be obtained from these systems. To further characterize these systems, the coordination number of the oxygens (both those on carbonate and those on water) around each $Gdm^+$ ion was calculated (Table 1).

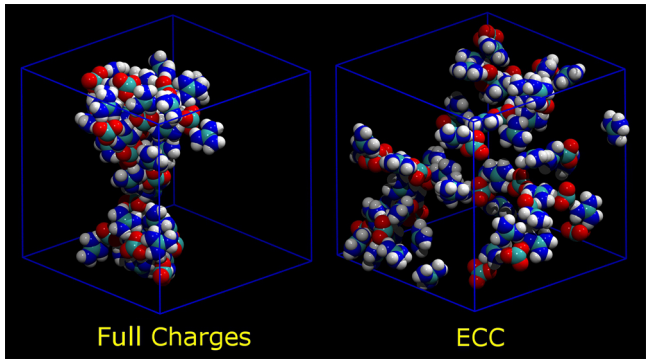



**Figure 1.** Snapshot of the two 1.5 m $Gdm_2CO_3$ solutions simulated in this study. The left image contains the full charges and has the ions almost entirely clustered, while the right has the electronic continuum correction model and has far less ion pairing. For clarity, water molecules are not shown.

As with previous studies of aqueous solutions of $K_2CO_3$,[21] the change in the ion−ion behavior upon the implementation

**Table 1. Coordination Number of Oxygen on Water ($O_w$) and Oxygen on Carbonate ($O_{carb}$) within 3.5 Å of an Atom on Guanidinium**

| | full charges | ECC |
|---|---|---|
| coordination number $O_{carb}$ | 4.8 | 2.3 |
| coordination number $O_w$ | 3.7 | 6.4 |
| coordination number $O_{total}$ | 8.5 | 8.7 |


dx.doi.org/10.1021/jp310719g | J. Phys. Chem. B 2013, 117, 1844−1848

of the ECC model (versus the full charges model) is dramatic (Figure 1).

In the charges simulation, the ions are practically completely ion paired into a single cluster with a lifetime in the 10−100s of nanoseconds range (which is very similar to previous MD studies of $Gdm_2CO_3$),[7] while in the ECC simulation, the fraction of ion pairing is significantly lower. This observation is further confirmed by the examination of the number of oxygens within 3.5 Å of any N on the guanidinium ion (Table 1). The total coordination numbers of oxygens for the full charge and ECC models are 8.5 and 8.7, respectively, with the former having a slightly lower coordination number due to the penetration of other species into this solvation shell. In the full charges model, 56% of this is due to oxygens on the carbonate (with the rest coming from water), even though the oxygens on carbonate only constitute ∼7.5% of the oxygens in the system. In the ECC model, only 26% of the oxygens in the first solvation shell are from the oxygens on the oxyanion. This factor of 2 in the coordination number of $O_{carb}$ has a significant effect on the bulk behavior of these ions. In the full charge simulation, the ions are almost entirely clustered, whereas in the ECC simulation, ion pairing and clustering is much weaker.

The density maps of carbonate oxygen around $Gdm^+$ also reveal this behavior (Figure 2). Not only is the coordination far more intense with the full charges simulations, but also, carbonate comes closer to $Gdm^+$ than in the ECC simulation. When both density maps are put on the same plot (Figure 2), the more extended yellow cloud from the full charge simulation thus penetrates closer to the $Gdm^+$ ion than the smaller red cloud from the ECC simulation.

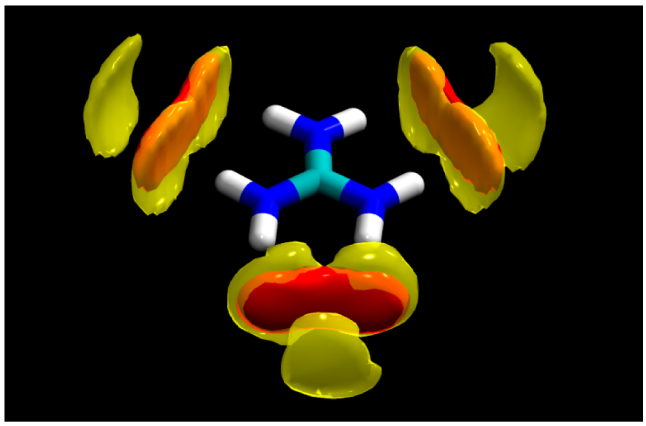


**Figure 2.** The density map of carbonate oxygen around $Gdm^+$. In both cases, the contour level is the same (0.07 atoms $Å^{-3}$, 31 times the bulk density of $O_{carb}$). The full charges simulation is shown in yellow, while the electronic continuum correction model is shown in red. Clearly, the full charges model has a higher level of ion pairing, and also, the approach of the two ions is closer than that in the ECC model.

## COMPARISON OF MOLECULAR DYNAMICS WITH NEUTRON SCATTERING

The experimental measurement from the previous study[7] constitutes the summation of three weighted pairwise correlations of the substituted nitrogens to the nitrogen, carbon, and oxygen in the system (Figure 3).

As can be seen, the function calculated for each respective simulation reflects the different ion−ion behavior of each system. This is visible in all three of the components (NN, NO, and NC). Further, it turns out to be a significant feature of these simulations that the molecular NN peak is slightly shifted from 2.25 for the full charges simulation to 2.30 for the ECC simulations. This change in position is mostly due to the stronger ion pairing in the full charges simulation distorting the molecular structure of the $Gdm^+$ ion. Similarly, the more intense ion pairing and clustering in the full charge simulation compared to that in the ECC simulation is clearly visible in the carbonate oxygen component of the $g_{NO}(r)$ term. This can be seen primarily in the relative heights of the peaks at ∼2.9 Å in Figure 3.

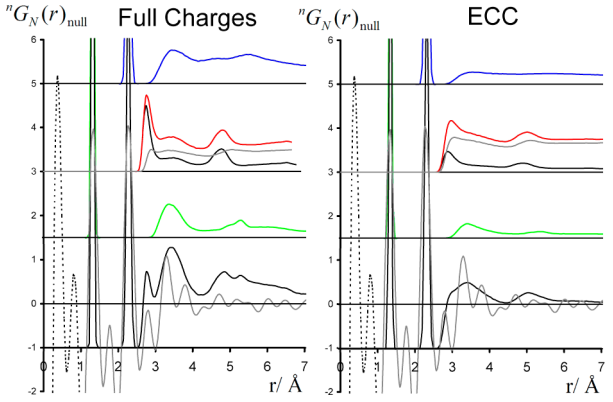



**Figure 3.** (Left) The components of the function $^nG_N(r)_{null}$ as calculated from the full charges simulation. (Right) The same functions for the electronic continuum correction model. In both cases, the color scheme is the same, with the upper (blue) being the NN component of the function $^nG_N(r)_{null}$, red being the NO component, and green being the NC component. In the case of the NO component, this is further broken down into the fraction of this due to the $NO_{carb}$ (black) and $NO_w$ (gray) component. The lower black shows the total function as calculated from the MD simulations. Gray shows the equivalent function from the experimental data.

To generate the experimental function from the simulation data, all three of these weighted pairwise distribution functions are summed. This can then be back-transformed to give the *Q*-space version of the data[19] (the one that was actually measured at the diffractometer[7]). These are shown for both the full charges and ECC simulation in Figure 4, demonstrating the much better agreement with neutron scattering for the latter.

The general relationship between *Q*-space and *r*-space data is that sharp intramolecular correlations (which tend to be at shorter *r*) yield longer wavelength features in *Q*-space. This can most easily be demonstrated by deleting the first peak in the experimental and MD *r*-space data (setting the function to the low *r* limit over this data range and then back-transforming both the MD and experimental data sets, middle panel in Figure 4). Direct comparisons between simulation and experiment are better performed in *Q*-space on several grounds. First, if there is a very sharp molecular correlation type peak in the *r*-space data, it will require an extended *Q*-range to fully describe it. If, however, that *Q*-range is not available in the experimental data, the feature will appear broader in the experimental *r*-space data than that in the MD *r*-space data. However, if the comparison is done in *Q*-space, the data should match fully for the entire *Q*-range. Second, many of the longer-range structures that we are mostly interested in for these solutions occur at higher *r* and are broader features. This means that if the comparison is performed in *r*-space, these broad features will be difficult to discern over the "ringing" noise, which tends to hamper the interpretation of such data. If,


dx.doi.org/10.1021/jp310719g | J. Phys. Chem. B 2013, 117, 1844−1848

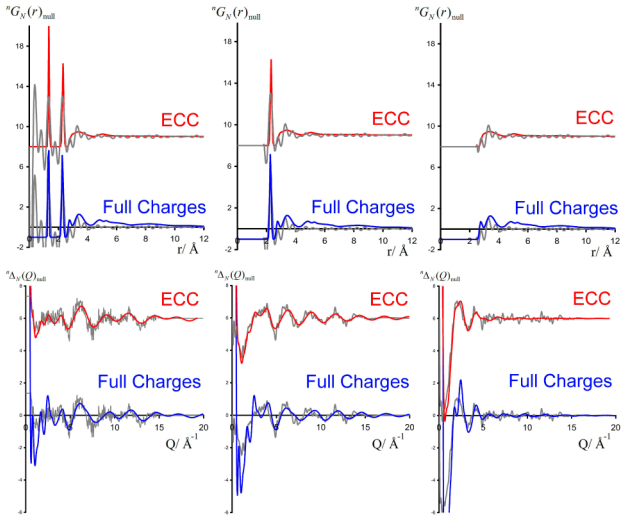



**Figure 4.** The upper set of graphs is the real space function ($^{n}G_N(r)_{null}$), and the lower are the *Q*-space functions ($^{n}\Delta_N(Q)_{null}$). In each case, the experimental data are shown in gray, while the functions calculated from the MD are shown in color (blue for full charges and red for the electronic continuum correction model). The data shown in the top left panel is the same as that shown in Figure 3. The left panel contains the full data range, the middle panel has been set to the low *r* limit in the range of 0−1.5 Å, and the right panel is in the range of 0−2.7 Å. The lower panels are the back-transforms of the upper panels data. The two sharp peaks in $^{n}G_N(r)_{null}$ (top left) at 1.5 and 2.3 Å, respectively, are due to the CN and NCN intramolecular correlations.

however, the comparison is done in *Q*-space, then this longer-range structure will manifest itself as a sharper "high-frequency" feature at lower *Q* (typically below 5 $A^{-1}$ for most hydration-type structures). For this reason, the *Q*-space data are of greatest interest in the *Q*-range of 0−5 $Å^{-1}$, while the higher *Q*-region mostly contains information on molecular motifs that are of little interest (the CN and NCN molecular correlations of the guanidinium ion).

After the deletion of the first CN peak at ∼1.4 Å in both MD and experimental data, the discrepancies in the comparison of the full charges with the experimental *Q*-space data become more evident (lower panels of Figure 4). It is clear that the experimental lower *Q* data region (which mostly corresponds to broader, higher *r* structures such as the hydration of the ion) is fitted much better with the ECC data than with the full charges data. On top of this, the fit of the data in the range of 5−10 $Å^{-1}$ is also much better with the ECC data than that with the full charges data. This can be shown to be due to the second molecular peak in the function $^{n}G_N(r)_{null}$. When the first intramolecular peak (due to the NC bond) at 1.5 Å is deleted (the function is set to the low *r* limit in the range of 0−1.5 Å) and the data are back-transformed into *Q*-space, the discrepancy between the full charges simulation and the experimental data remains (middle panels of Figure 4). However, when this process is repeated with the intramolecular guanidinium peak at 2.3 Å (due to the NCN motif), the discrepancy in the *Q*-space data is no longer found. The distortion of the $Gdm^+$ ion due to the strong ion pairing in the full charges simulation is less consistent with the experimental data than the practically undistorted $Gdm^+$ from the ECC simulation.

Arguably, the simplest way to see these small differences in the intramolecular structure of these two solutions (top left panel of Figure 4) is by comparing the relative heights of the two molecular correlations (the CN and NN correlations) of both the full charges and ECC simulations. Given that the integration of these two peaks must be the same for both simulations (due to the same molecular topology), the mere fact that the peaks are of different heights for the two simulations demonstrates that there must be different intramolecular configurations. Further, after all of the molecular correlations have been deleted (farthest right panel of Figure 4), when these data are back-transformed, it is found that the fit of the ECC data to the experiment is again much better in the range of 0−5 $Å^{-1}$ compared to the fit of the full charges simulation. It is this improvement in the fitting of the lower *Q* data in all three comparisons that is the most robust indicator that the ECC model is performing significantly better in emulating the structure of the experimental aqueous guanidinium carbonate solution than the full charges model. This also suggests that in full charge, nonpolarizable simulations, the ion−protein interactions, such as those found between the guanidinium side chains of arginine and the carbonate ions in proteins such as lactoferrins, may be overestimated.

## CONCLUSIONS

Simulations were performed for aqueous $Gdm_2CO_3$ using full charges, as well as scaled charges within the ECC approach, which implicitly accounts for electronic polarization effects. The predictions of the structure from the simulations were then compared to previous neutrons scattering measurements of the same system.[7] This comparison suggests that the ion−ion interactions are significantly stronger in the full charges simulations, leading to a more extensive ion pairing and clustering than those in the ECC model. It also indicates that the ECC model is performing significantly better than the full charges model, emulating quantitatively the structure obtained via the experimental measurement. This result is also suggestive for similar ion pairing structures found in proteins,[2,22] which may be overemphasized in full charges simulations in nonpolarizable water and where ECC may represent a simple way to improve the results.

## ASSOCIATED CONTENT

### Supporting Information

A video summary of the paper. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

### Corresponding Author

*E-mail: philip.mason@uochb.cas.cz.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Support from the Czech Science Foundation (Grant P208/12/G016) is gratefully acknowledged. P.J. thanks the Academy of Sciences for the Praemium Academie award.

## REFERENCES

(1) (a) Puglisi, J. D.; Chen, L.; Frankel, A. D.; Williamson, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90* (8), 3680−3684. (b) Rothbard, J. B.; Kreider, E.; Vandeusen, C. L.; Wright, L.; Wylie, B. L.; Wender, P. A. *J. Med. Chem.* **2002**, *45* (17), 3612−3618.

(2) Schneider, C. P.; Shukla, D.; Trout, B. L. *J. Phys. Chem. B* **2011**, *115* (22), 7447−7458.
(3) (a) Scholtz, J. M.; Qian, H.; Robbins, V. H.; Baldwin, R. L. *Biochemistry* **1993**, *32* (37), 9668−9676. (b) Olson, C. A.; Spek, E. J.; Shi, Z. S.; Vologodskii, A.; Kallenbach, N. R. *Proteins: Struct., Funct., Genet.* **2001**, *44* (2), 123−132.
(4) Baker, H. M.; Anderson, B. F.; Baker, E. N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (7), 3579−3583.
(5) Ward, P. P.; Paz, E.; Conneely, O. M. *Cell. Mol. Life Sci.* **2005**, *62* (22), 2540−2548.
(6) Adams, T. E.; Mason, A. B.; He, Q. Y.; Halbrooks, P. J.; Briggs, S. K.; Smith, V. C.; MacGillivray, R. T. A.; Everse, S. J. *J. Biol. Chem.* **2003**, *278* (8), 6027−6033.
(7) Mason, P. E.; Neilson, G. W.; Kline, S. R.; Dempsey, C. E.; Brady, J. W. *J. Phys. Chem. B* **2006**, *110* (27), 13477−13483.
(8) Leontyev, I. V.; Stuchebrukhov, A. A. *J. Chem. Theory Comput.* **2010**, *6* (5), 1498−1508.
(9) Schroder, C. *Phys. Chem. Chem. Phys.* **2012**, *14* (9), 3089−3102.
(10) Leontyev, I.; Stuchebrukhov, A. *Phys. Chem. Chem. Phys.* **2011**, *13* (7), 2613−2626.
(11) Mason, P. E.; Neilson, G. W.; Enderby, J. E.; Saboungi, M. L.; Dempsey, C. E.; MacKerell, A. D.; Brady, J. W. *J. Am. Chem. Soc.* **2004**, *126* (37), 11462−11470.
(12) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91* (24), 6269−6271.
(13) Leontyev, I. V.; Stuchebrukhov, A. A. *J. Chem. Theory Comput.* **2010**, *6* (10), 3153−3161.
(14) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103* (19), 8577−8593.
(15) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81* (8), 3684−3690.
(16) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34* (5), 1311−1327.
(17) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. In *AMBER 11*; University of California: San Francisco, CA, 2011.
(18) Beauchamp, K. A.; Lin, Y. S.; Das, R.; Pande, V. S. *J. Chem. Theory Comput.* **2012**, *8* (4), 1409−1414.
(19) Mason, P. E.; Neilson, G. W.; Dempsey, C. E.; Brady, J. W. *J. Am. Chem. Soc.* **2006**, *128* (47), 15136−15144.
(20) Mason, P. E.; Dempsey, C. E.; Neilson, G. W.; Brady, J. W. *J. Phys. Chem. B* **2005**, *109* (50), 24185−24196.
(21) Mason, P. E.; Wernersson, E.; Jungwirth, P. *J. Phys. Chem. B* **2012**, *116* (28), 8145−8153.
(22) (a) Shukla, D.; Schneider, C. P.; Trout, B. L. *J. Am. Chem. Soc.* **2011**, *133* (46), 18713−18718. (b) McLain, S. E.; Soper, A. K.; Daidone, I.; Smith, J. C.; Watts, A. *Angew. Chem., Int. Ed.* **2008**, *47* (47), 9059−9062.