Large-scale first-principles molecular dynamics simulations: application to the microsolvation of biologically-relevant ions in aqueous clusters

Chun C Mak, Denise M Koch and Gilles H Peslherbe*

Centre for Research in Molecular Modeling and Department of Chemistry and Biochemistry, Concordia University, 7141 Sherbrooke St. W., Montréal, Québec, Canada, H4B 1R6

*gilles.peslherbe@concordia.ca

Abstract. First-principles molecular dynamics simulations are employed to investigate the solvation structure of the biologically-relevant paradigm guanidinium-ion in water clusters, Gdn\(^{+}(H_2O)_n\). In these simulations, the intermolecular interactions are evaluated along the trajectories with the self-consistent-charge density-functional tight-binding (SCC-DFTB) model with the on-site third-order correction and modified effective Coulomb interaction, an approximate model of density-functional theory. For Gdn\(^{+}(H_2O)_{21}\), the simulated probability distribution of the ion-to-water cluster centre-of-mass distances and water angular coordinates suggests that Gdn\(^{+}\) is primarily localised at the surface of the water cluster, with the surface of the ion relatively devoid of water molecules. Estimates of the computational resources required for first-principles molecular dynamics simulations of larger Gdn\(^{+}(H_2O)_n\) (n≤100) indicate that, with a modern supercomputer, such simulations can readily be performed in a matter of days.

1. Introduction

Proteins play numerous important roles in biology, serving for instance as enzymes to catalyze vital chemical reactions, transporters of chemical substances between different parts of an organism, and as regulators of biological processes. The function of a given protein is closely related to its conformation, and alterations to its native conformation may result in loss of function and in disease. It has been recognised for some time that molecular interactions involving the guanidinium ends of the side chains of arginine residues may play a crucial role in the determination of the structure of proteins and protein complexes [1]. Therefore, a molecular-level understanding of such interactions may be necessary to develop a clearer picture of the determinants of structure in biomolecules.

The guanidinium ion, \([\text{C(NH}_2)_3]^+\) or Gdn\(^{+}\), is an excellent model of the positively-charged guanidinium group of arginine and, since Gdn\(^{+}\) is widely believed to be responsible for the activity of several powerful protein denaturant salts, much experimental and theoretical work has been carried out to understand its hydration structure and its interaction with other dissolved solutes in aqueous solutions [2-4]. The general picture arising from previous work on aqueous solutions of Gdn\(^{+}\) is that Gdn\(^{+}\) is very weakly hydrated and it exhibits a high propensity to aggregate with itself, other ions, as well as small aromatic molecules; the latter may have implications for the role of the Gdn\(^{+}\) group in protein folding, protein-protein interactions and protein denaturation [2]. The existence of these ionic
aggregates in aqueous Gdn\(^+\) solutions has not been undisputed [3]; however, calculations of the potential of mean force of association of Gdn\(^+\) with itself or with small aromatic molecules in aqueous solutions generally suggest that such aggregates are thermodynamically stable [4].

The investigation of gas-phase guanidinium-water clusters of finite size, Gdn\(^+(\text{H}_2\text{O})_n\), may offer a complementary perspective on the hydration of Gdn\(^+\), which could potentially result in a more in-depth understanding of the seemingly unusual behaviour of Gdn\(^+\) in aqueous media. On the experimental side, Hof and co-workers have used energy-dependent electro spray ionisation mass spectrometry to investigate the dehydration process of Gdn\(^+(\text{H}_2\text{O})_n\) and various methylated derivatives Me\(_m\)Gdn\(^+(\text{H}_2\text{O})_n\) (m=0-4 and 6; n=1-50) [5]. Their work demonstrated that methylation of Gdn\(^+\) results in increased hydrophobicity of the ion, highlighting the possible role of hydration/dehydration in the modulation of biologically-relevant intermolecular interactions involving the Gdn\(^+\) group. In this contribution, the use of reliable, first-principles molecular dynamics simulations to investigate the hydration of Gdn\(^+\) in aqueous clusters is first described, and the approach is applied to characterise the solvation structure of small Gdn\(^+(\text{H}_2\text{O})_n\). It will then be shown that, with a modern supercomputer, an investigation of large Gdn\(^+(\text{H}_2\text{O})_n\) (n<100) using the same first-principles molecular dynamics approach is readily attainable.

2. Computational Procedure

Molecular dynamics simulations were used to investigate the structural properties of Gdn\(^+(\text{H}_2\text{O})_n\) and determine the preferential position of Gdn\(^+\) within the cluster. Two coordinates, defined in figure 1, the ion-to-water cluster centre-of-mass separation \(r_{\text{cm}}\) and the angles \(\theta\) between a given water molecule, the centre of mass of Gdn\(^+\) and the centre of mass of the water cluster were used to discriminate between the predominant solvation structures, interior or surface, of Gdn\(^+(\text{H}_2\text{O})_n\) as proposed in earlier work [6].

![Figure 1. Definition of the ion-to-water cluster centre-of-mass distance, \(r_{\text{cm}}\), and \(\theta\), the angle between a given water molecule, the ion and the water cluster centres of mass.](image)

In order to ensure adequate sampling of Gdn\(^+(\text{H}_2\text{O})_n\) configurations, including interior and surface-solvated cluster structures, the umbrella sampling method was employed [7]. In this method, a series of molecular dynamics simulations are performed, each in the presence of an added umbrella potential \(U_{\text{umb}}\) centred on a different value of \(r_{\text{cm}}\) (denoted \(r_{\text{cm}}^0\)) to restrain the ion-to-water cluster centre-of-mass separation to values near \(r_{\text{cm}}^0\).

\[
U_{\text{umb}} = \frac{1}{2}k(r_{\text{cm}} - r_{\text{cm}}^0)^2
\]

For Gdn\(^+(\text{H}_2\text{O})_{21}\), the value of the force constant \(k\) of the umbrella potential was chosen to be 7.0 kcal-mol\(^{-1}\)Å\(^{-2}\), while the \(r_{\text{cm}}^0\) values ranged from 0.0 Å to 8.0 Å, with each value spaced by 1.0 Å; this prescription ensured efficient generation of a smooth potential of mean force for this cluster size. Each molecular dynamics simulation was run for 2 nanoseconds using the velocity Verlet algorithm [8], with 1 nanosecond of equilibration followed by an additional nanosecond of simulation for data collection. The energies and energy gradients of Gdn\(^+(\text{H}_2\text{O})_n\) used in the molecular dynamics
simulations were computed using approximate density-functional theory, namely the self-consistent-charge density-functional tight-binding (SCC-DFTB) model with the on-site third-order correction and a modified effective Coulomb interaction [9]. This model, which is a modification of the traditional SCC-DFTB model to provide an improved description of hydrogen-bonding interactions and proton affinities, appears to adequately describe the relative strength of Gdn\(^+\)-H\(_2\)O and H\(_2\)O-H\(_2\)O interactions, a crucial requirement for reliable simulation results [10]. The approximate density-functional theory calculations were performed with the DFTB+ computer code [11]. A Nosé-Hoover thermostat was applied to maintain the temperature of the system at 298 K [12]. Furthermore, to prevent the escape of water molecules from Gdn\(^+\)(H\(_2\)O)\(_n\) and thus ensure that a proper canonical ensemble of cluster configurations is sampled, a restraining potential \(U_{\text{rstr}}\) was added whenever the distance \(d\) between any given water molecule and the centre of mass of the water cluster exceeded \(d_{\text{max}}\).

\[
U_{\text{rstr}} = \frac{1}{2} K (d - d_{\text{max}})^2
\]

The optimal values of \(K\) and \(d_{\text{max}}\) are dependent on the cluster size \(n\), and in general, are chosen such that the cluster geometries are only slightly affected by the presence of \(U_{\text{rstr}}\). For Gdn\(^+\)(H\(_2\)O)\(_{21}\), optimal values of \(K\) and \(d_{\text{max}}\) were found to be 200 kcal·mol\(^{-1}\)Å\(^{-2}\) and 15 Å, respectively.

The calculation of the potential of mean force for Gdn\(^+\)(H\(_2\)O)\(_n\) requires energy and gradient calculations for a total of \(1.8 \times 10^7\) cluster configurations, which represents a significant investment of computational resources. Preliminary test simulations were performed to determine the optimal distribution of the required calculations on a cluster of Intel® Xeon® X5560 processors consisting of nodes that each have 8 cores. Results of these preliminary simulations, shown in figure 2, indicate that the elapsed time for a given simulation decreases dramatically as the number of cores used is increased from one to four, with further significant decreases up to 8 cores.

![Figure 2. Real time elapsed for 2-nanosecond molecular dynamics simulations of Gdn\(^+\)(H\(_2\)O)\(_{21}\) using different numbers of cores](image)

The simulations for each value of \(r_{\text{cm}}^0\) were therefore distributed on distinct nodes of the computer cluster, with 8 cores used for every simulation. All 9 independent simulations were carried out simultaneously, minimising the elapsed time needed for the entire set of simulations, which were completed in around 10 days, requiring a total of 17280 cpuxhours (the equivalent of 2 years on one core processor). Once the simulations were completed, the weighted histogram analysis method (WHAM) was used to remove the effects of the biasing potential on the probability distribution of sampled configurations and to reconstruct the complete probability distribution of \(r_{\text{cm}}\) values from the 9 individual simulations [13].
3. Results and Discussion
The probability distribution $P(r_{cm}, \theta)$ for Gdn$^+(H_2O)_{21}$ obtained from the molecular dynamics simulations is shown in figure 3; it exhibits one single peak at $r_{cm} = 4$ Å, $\theta \approx 30^\circ$. For a spherical cluster of uniform density in which the Gdn$^+$ is internally solvated, $P(r_{cm}, \theta)$ should peak at $r_{cm} \approx 0$ Å and exhibit a dependence on $\theta$ proportional to $\sin(\theta)$ i.e. an isotropic angular distribution. By contrast, $P(r_{cm}, \theta)$ for similar spherical clusters in which the ion is located at the surface would exhibit a peak at values of $r_{cm}$ that approach the radius of the cluster, with very low $P(r_{cm}, \theta)$ at values of $\theta$ larger than $90^\circ$. Since $P(r_{cm}, \theta)$ for Gdn$^+(H_2O)_{21}$ peaks at approximately $r_{cm} = 4$ Å, which is about half the average radius (around 8 Å) of the water cluster, it is difficult to draw firm conclusions on the preferential solvation structure of Gdn$^+(H_2O)_{21}$ from the most probable value of $r_{cm}$ alone. The plot of $P(r_{cm}, \theta)$ indicates that there is a notable absence of water molecules located at angles $\theta$ larger than $90^\circ$ for all values of $r_{cm}$ for which the value of $P(r_{cm}, \theta)$ is non-negligible, indicating that the water molecules are not evenly distributed about Gdn$^+$. Typical Gdn$^+(H_2O)_{21}$ configurations at $r_{cm} = 4$ Å are shown in figure 4. Despite the fact that $r_{cm}$ is much smaller than the radius of the cluster, Gdn$^+$ is obviously located at the surface due to the non-spherical geometry of these Gdn$^+(H_2O)_{21}$ configurations. The
surfaces of Gdn⁺ appear to be relatively devoid of water molecules in these configurations, with most Gdn⁺-H₂O interactions involving the NH₂ groups of the ion. Figure 5 shows two different views of the average spatial distribution of the water molecules around Gdn⁺ in Gdn⁺(H₂O)₂¹, providing a more complete picture of Gdn⁺ hydration in small aqueous clusters. These plots were generated by superposing the Gdn⁺(H₂O)₂¹ configurations sampled from the molecular dynamics simulation centred at rₓₐₜ = 4 Å, with Gdn⁺ fixed in space and identically oriented in all configurations. Hydrogen atoms are omitted for clarity. Regions around Gdn⁺ with a heavier density of points correspond to more hydrated regions i.e. regions of the cluster where water molecules are more likely to be located. These plots show that there is a higher probability for water molecules to be located in three “islands” around Gdn⁺ outwards from the “pockets” defined by adjacent NH₂ groups, suggesting that there may be an increased propensity for water molecules to interact with these groups. There is, on the other hand, a lower probability for water molecules to be found on the surfaces of Gdn⁺, a possible reflection of the seemingly hydrophobic nature of Gdn⁺ in certain environments. The results from these reliable first-principles molecular dynamics simulations of Gdn⁺(H₂O)₂¹ corroborate the findings of previous simulations of Gdn⁺ and its salts in aqueous solution using various model potentials, which indicate that Gdn⁺ is extremely weakly hydrated with surfaces relatively devoid of water molecules [2]. These findings may have important implications for understanding interactions involving Gdn⁺ groups in biological systems, where the dehydrated face of the planar ion may be available for binding with non-polar molecules.

While the present findings already shed new light onto the hydration of Gdn⁺, a thorough theoretical investigation of Gdn⁺(H₂O)ₙ for various cluster sizes may be necessary to develop a sound understanding of the role of hydration in the modulation of biologically-relevant interactions involving Gdn⁺ groups. With powerful, modern supercomputers, the first-principles molecular dynamics approach employed in the present work to investigate the solvation structure of Gdn⁺(H₂O)₂¹ is readily applicable to larger Gdn⁺(H₂O)ₙ (n≤100). Preliminary molecular dynamics simulations of various Gdn⁺(H₂O)ₙ (2≤n≤100) using the SCC-DFTB model with the on-site third-order correction and modified effective Coulomb interaction have been performed to assess the feasibility of using large-scale first-principles molecular dynamics simulations to investigate the solvation structure of larger clusters. For each cluster size, a single simulation was run on one node of a cluster of Intel® Xeon® X5560 processors using 8 cores with the wall-time limit set to 8 hours, and the number of simulation steps completed within this time was determined. The data were then used to estimate the time
required to complete a 2-nanosecond simulation for each Gdn\(^+\)(H\(_2\)O)\(_n\), and the results are summarised in figure 6. The time required to complete a single 2-nanosecond simulation increases almost linearly with cluster size, with one simulation requiring 10 days for Gdn\(^+\)(H\(_2\)O)\(_{20}\) and the same run requiring 50 days for Gdn\(^+\)(H\(_2\)O)\(_{100}\). In all fairness, the computational cost of the potential of mean force calculation will not scale linearly with cluster size. Firstly, a larger number of umbrella sampling windows might be needed as the cluster grows in size and its radius becomes larger. Secondly, a larger number of configurations might need to be sampled as the density of cluster states increases along with cluster size, a point that is, however, not grounded on a solid theoretical foundation. Assuming that a complete P(r\(_{cm}\), \(\theta\)) could be obtained using the WHAM method with data from 12 independent simulations restrained around different values of r\(_{cm}\), the solvation structure of Gdn\(^+\)(H\(_2\)O)\(_{100}\) could be characterised in 50 days under ideal conditions in which uninterrupted calculations are possible on 12 8-core compute nodes. Such a simulation would entail energy and gradient calculations for a total of 2.4\times10\(^{10}\) cluster configurations and require the use of \(\sim115200\) cpu\times hours (\(\sim13\) years of a single core processor) of computing time, which while demanding is still feasible on most modern supercomputers. Finally, as the number of cores per node tends to increase dramatically on modern clusters, we anticipate the real-time cost of the calculations to decrease even further on new cluster infrastructures.

4. Summary and Conclusions
In this contribution, first-principles molecular dynamics simulations using the SCC-DFTB model with the on-site third-order correction and modified effective Coulomb interaction, which reliably describes the relative strength of the Gdn\(^+\)-H\(_2\)O and H\(_2\)O-H\(_2\)O interactions when compared with high-level \textit{ab initio} electronic structure theory calculations, were used to investigate the solvation structure of Gdn\(^+\)(H\(_2\)O)\(_{21}\). The simulation results tentatively indicate that, in small Gdn\(^+\)(H\(_2\)O)\(_n\), Gdn\(^+\) is preferentially located at the surface of the water cluster, with the ion surface essentially dehydrated. These preliminary findings, which are consistent with the conclusions of previous work on the hydration structure of Gdn\(^+\) in aqueous solutions, provide complementary, molecular-level insight into the apparent “hydrophobicity” of Gdn\(^+\). The feasibility of applying first-principles molecular dynamics simulations to the investigation of larger Gdn\(^+\)(H\(_2\)O)\(_n\) was explored, laying the groundwork for a comparative investigation of Gdn\(^+\) hydration in aqueous clusters of varying sizes that may lead to a more in-depth understanding of the influence of hydration on biologically important intermolecular interactions involving Gdn\(^+\) groups.

Acknowledgements
This work was funded by research grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada. C.C.M. is the recipient of an NSERC award. Computational resources
were provided by the Réseau Québécois de Calcul Haute Performance (RQCHP) and the Centre for Research in Molecular Modeling (CERMM), which was established with the financial support of the Concordia University Faculty of Arts & Science, the Ministère de l’Education du Québec (MEQ) and the Canada Foundation for Innovation (CFI). G.H.P. holds a Concordia University Research Chair.

References

[1] Pednekar D, Tendulkar A and Durani S 2009 Proteins 74 155
Crowley P and Golovin A 2005 Proteins 59 231
Magalhaes A, Maigret B, Hoflack J, Gomes J and Scheraga H 1994 J. Protein Chem. 13 195

[2] Mason P, Neilson G, Dempsey C, Barnes A and Cruickshank J 2003 Proc. Natl. Aca. Sci. USA 100 4557
Mason P, Dempsey C, Neilson G, Kline S and Brady J 2009 J. Am. Chem. Soc. 131 16689
Mason P, Neilson G, Kline S, Dempsey C and Brady J 2006 J. Phys. Chem. B 110 13477
Mason P, Dempsey C, Neilson G and Brady J 2005 J. Phys. Chem. B 109 24185
Mountain R and Thirumalai D 2004 J. Phys. Chem. B 108 19711
Mason P, Neilson G, Enderby J, Saboungi M, Dempsey C, Mackerell A and Brady J 2004 J. Am. Chem. Soc. 126 11462

[3] Hunger J, Niedermayer S, Buchner R and Hefter G 2010 J. Phys. Chem. B 114 13617

[4] Soetens J, Millot C, Chipot C, Jansen G, Ángyán J and Maigret B 1997 J. Phys. Chem. B 101 10910
Saigal S and Pranata J 1997 Bioorg. Chem. 25 11
Duffy E, Kowalczyk P and Jorgensen W 1993 J. Am. Chem. Soc. 115 9271
Boudon S, Wipff G and Maigret 1990 J. Phys. Chem. 94 6056

[5] McQuinn K, McIndoe J and Hof F 2008 Chem. Eur. J. 14 6483

[6] Peslherbe G, Ladanyi B and Hynes J 2000 Chem. Phys. 258 201

[7] Torrie G and Valleau J 1974 Chem. Phys. Lett. 28 578

[8] Swope W, Andersen H, Berens P and Wilson K 1982 J. Chem. Phys. 76 637

[9] Yang Y, Yu H, York D, Cui Q and Elstner M 2007 J. Phys. Chem. A 111 10861
Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S and Seifert G 1998 Phys. Rev. B 58 7260

[10] Mak C, Koch D and Peslherbe G Article to be submitted

[11] Aradi B, Hourahine B and Frauenheim T 2007 J. Phys. Chem. A 111 5678

[12] Nose S 1984 J. Chem. Phys. 81 511
Hoover G 1985 Phys. Rev. A 31 1695

[13] Kumar S, Bouzida D, Swendsen R, Kollman P and Rosenberg J 1992 J. Comput. Chem. 13 101