J. Am. Chem. Soc. **1994,** *116,* 7451–7452

Grand Canonical Monte Carlo Simulation of Water Positions in Crystal Hydrates

Haluk Resat and Mihaly Mezei*

Department of Biophysics and Physiology
Mount Sinai School of Medicine, CUNY
New York, New York 10029-6574

Received May 9, 1994

X-ray crystallography is an important experimental method to investigate chemical structures, proving to be particularly powerful for studying the structure and function of biological macromolecular systems.[1] For large molecules, however, the resolution of the data limits the accuracy of the structure obtained. Perhaps the most challenging part of the structure refinement process is the determination of both the location and the number of solvent (water) molecules. This report proposes the use of the cavity-biased grand canonical Monte Carlo simulation to address the problem of accurate determination of the water positions in crystal structures, and it communicates our initial results for a particular crystal hydrate system.

As our test crystal system, the orthorhombic form of the sodium salt of hyaluronic acid (HYA) was chosen. HYA is a linear polydisaccharide of the form (-GCU-NAG-)$_n$ where GCU is glucuronic acid and NAG is *N*-acetylglucosamine.[2] The crystal structure and the unit cell parameters of HYA were taken from the Protein Data Bank (entry, 3HYA; resolution, 3 Å). Each HYA unit cell comprises four asymmetric subunits having a $P2_12_12_1$ space-group symmetry and consists of two antiparallel left-handed helix forming chains. In the unit cell, there are eight residues of each of GCU and NAG, eight sodium ions, and 16 experimentally determined water molecules.[2] Each sodium ion is in a 6-fold cage formed by the oxygens of different HYA chains and by water molecules, Figure 1. In terms of the surrounding environment and hydrogen-bonding mechanisms, the rest being symmetry related, only two of the sodium ions and four of the water molecules are distinct.

The interaction parameters of HYA were represented by the AMBER force field,[3] and the TIP3P model[4] was used for water. HYA site charges were generated using *ab initio* methods.[5] The calculated charges for both distinct sodiums were almost equal, and therefore, the same value, 0.95, was assigned to both. The cavity-biased grand canonical Monte Carlo (CB-GCMC) technique[6] with a one to one ratio of insertion/deletion to regular move attempt was employed. Since our main aim was to determine the mean water locations, the saccharide chains and the sodium ions were kept fixed in their crystal positions during the simulation; *i.e.*, only the water molecules were allowed to move.[7] After sufficient equilibration the simulations were run for five million steps each of regular and insertion/deletion moves. Of this, every other 500th configuration was collected for results analysis.

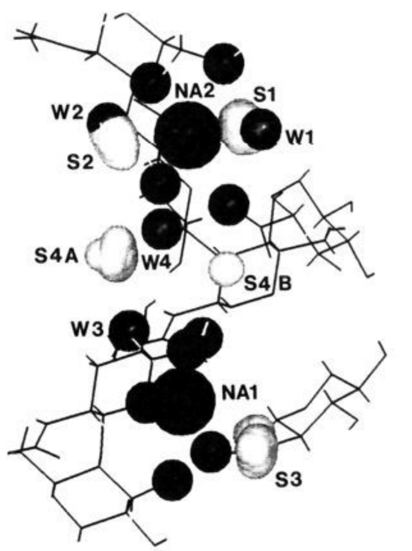

Figure 1. Structure around sodium ions in HYA. For clarity, uninvolved HYA residues are not shown. The labeling convention is that of ref 2. W1–4 (dark) are the experimentally reported waters, and S1–4 (gray) show the corresponding simulated locations highly occupied by the water oxygens. Unlabeled spheres are the saccharide oxygens bound to the sodiums. Spheres have half of their respective van der Waals sizes.

Use of the grand canonical ensemble to study the properties of crystal hydrates is essential for a couple of reasons: (a) The density measurement of the crystal is not accurate enough to measure the exact number of water molecules inside the unit cell.[2] Therefore, there may be considerable uncertainty in the number of hydrating water molecules. Use of the grand canonical ensemble partially solves this problem by allowing the number of water molecules to fluctuate. Thus, it can provide the statistical distribution profile and information of the energetics and the locations of added or removed water molecules at different configurations. (b) The water molecules of the crystal hydrates are mostly enclosed in disjoint pockets. An appropriate simulation technique has to allow for molecule exchange between the pockets. Conventional simulation methods, such as canonical and microcanonical ensemble simulations, fix the total number of water molecules and their distribution among the pockets from the start, and molecule exchange between the pockets is virtually impossible, resulting in a strong bias toward the initial configuration. MC simulations with very large step sizes would increase the frequency of the attempted particle exchange between the pockets, but in that case the move acceptance rate becomes very low, thus making it an impractical approach. In contrast, in the grand canonical ensemble, through molecule insertion/deletions, the water molecules are exchanged between the pockets. In addition, the cavity-biased formulation of GCMC was used because, at high densities, it considerably improves the sampling efficiency.[6]

Simulation results were analyzed using two different approaches, which resulted in identical conclusions. In the first, the solvent number density profile on a three-dimensional grid was calculated. The second approach utilized the generic solvent

---

(1) See, for example, Ringe, D.; Petsko, G. A. *Prog. Biophys. Mol. Biol.* **1985**, *45*, 197–235.

(2) Guss, J. M.; Hukins, D. W. L.; Smith, P. J. C.; Winter, W. T.; Arnott, S.; Moorhouse, R.; Rees, D. A. *J. Mol. Biol.* **1975**, *95*, 359–384.

(3) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profate, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.

(4) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926–935.

(5) *Ab initio* calculations with STO-3G basis sets (using the Gaussian-92 program, ref 11) were performed to generate the electrostatic potentials around the different subsections of HYA. The partial site charges were assigned by using the Merz–Kollman fitting procedure to the generated electrostatic potential. Respective subsections of HYA were defined as the two upchain and the two downchain residues surrounding each distinct sodium cation, and the water molecules in their experimentally determined positions which are involved in the cage forming around the sodiums.

(6) Mezei, M. *Mol. Phys.* **1980**, *40*, 901–906; **1987**, *61*, 565–582; **1989**, *67*, 1207–1208 (erratum). Note: In these references, the CB-GCMC method was denoted as CB/TVM.

(7) Within this treatment the utilized ensemble is not truly grand canonical, because it is a multicomponent system in which some components (HYA) are kept at constant number density and some (waters) at constant chemical potential. However, the label grand canonical ensemble is still appropriate due to the following reason. There is no free space to accommodate a second HYA molecule, and deleting the existing one would cost an enormous amount of energy (with minimum image cutoff, the water–HYA interaction energy is $\sim -300k_BT$). For this reason, even if allowed to fluctuate, the number of HYA molecules would hardly change, so the employed ensemble is essentially equivalent to a grand canonical ensemble.

0002-7863/94/1516-7451$04.50/0

site (GSS) idea, and the Hungarian method was used to determine the GSSs by overlapping the subsequent configurations.[8] In this approach, the particles do not carry labels, which allows for molecule interchange between the GSSs during the molecule assignment. The fact that the unit cell comprises four symmetry-related subunits can be used to investigate the convergence of the simulation runs. For this, using the mean locations calculated with the Hungarian method, the water molecules were partitioned into symmetry-related subgroups. The mean square deviation of symmetry-related GSSs from their averaged mean position was only 0.04 Å^2, implying a very well converged simulation run.

Comparison of the simulation and the experimental results for the mean water oxygen positions, Figure 1, shows that there is very good agreement for two of the water molecules, W1 and W2 (naming convention is that of ref 2). For these two water molecules, the relative displacements between the averaged simulation and the experimental positions are less than 1 Å. Even though there is a relatively large displacement, ~2.6 Å, a third water, W4, is also well reproduced, and both the experimental and the simulation results correspond to chemically possible locations. In correspondence to W4, two separate locations, S4A and S4B, are actually detected in the simulations. The observed disagreement for the fourth distinct water molecule is quite considerable, both quantitatively and chemically. If the saccharide oxygens bound to sodium ion NA1 are considered to approximately form a plane, then the experimental (W3) and the simulation results (S3) are on different sides of that plane.

Due to the presence of the sodium ions and to the close-packed nature of the crystals, electrostatic interactions are expected to play an important role in the chemical structure of crystal hydrates. To test if the found disagreements are electrostatic in origin, we performed a second simulation using an approximate MC acceptance criterion[9] which downplays the importance of the electrostatic forces. In this second simulation, the S3 peak was again observed and it was as strong as before. In addition another peak (not shown in Figure 1) with considerable density was observed almost equidistant from W3 and W4, 3.8 Å. These findings imply that, even though electrostatic effects may be very important, they are probably not the main contributor to the observed discrepancies. This, of course, does not rule out the possibility that the point charge representation of the electrostatic interactions may have been inadequate and going to higher level presentation would be necessary. Other possible sources of discrepancy are the deficiencies in the potential parameter set used in the simulations, or the improper modeling in the experimental structure refinement analysis (note that the refined structure has an R-factor[2] of only 0.3). For example, the simulation results suggests that there may actually be five rather than four distinct water molecules (or four molecules with fractional occupancies at different sites), and the necessary empty volume in the crystal structure exists to accommodate this additional water. Note also that the sodium ions and HYA chains were not allowed to move in the simulation. The determination of the location of the sodium ions in X-ray crystallography is not that straightforward either; therefore, it is possible that the cations might have been slightly misplaced, which in turn may contribute to the observed disagreement. Investigation of the effects of using different interactions potential parametrization and of allowing for the movement of the sodium ions during the simulation is currently in progress. An interesting search would be to construct a structure model based on the simulation results reported here and compare with the experimental X-ray diffraction patterns to see if there is any improvement in the agreement.

This study clearly shows that the use of the grand canonical ensemble in crystal hydrate simulations overcomes many of the intrinsic difficulties encountered in the analogous canonical or microcanonical ensemble simulations and is able to reliably predict the water positions for a given potential function model. As its implementation in this report shows,[10] with its cavity-biased formulation, grand canonical Monte Carlo simulation is an effective and robust method and can give leads into a possible array of alternative structures to investigate. It is computationally not too expensive and can be rather easily combined into crystal structure refinement protocols to supplement model building. This would improve the accuracy of the refined structures by establishing an independent test so as to reveal the erroneously built structures. Or, in an alternative utilization, starting with an accurate crystal structure, the CB-GCMC simulation method can equivalently be used to test the potential function models.

Acknowledgment. This work was supported by NIH Grant No. R55-GM43500. Computing resources were provided in part by the University Computing Center of the City University of New York. We thank a referee for raising the point discussed in footnote 7 and Dr. Marc Glucksman for helpful discussions.

---

(8) Mezei, M.; Beveridge, D. L. *J. Comput. Chem.* **1984**, *5*, 523–527.

(9) Resat, H.; Mezei, M. In preparation. The approximate criterion ignores the electrostatic forces, and the acceptance of moves is determined by considering solely the short-range Lennard-Jones forces.

(10) For a related use of GCMC in zeolites/microporous solids, see: Jameson, C. J.; Jameson, A. K.; Baello, B. I.; Lim, H.-M. *J. Chem. Phys.* **1994**, *100*, 5965–5976. Van Tassel, P. R.; Davis, H. T.; McCormick, A. V. *J. Chem. Phys.* **1993**, *98*, 8919–8928.

(11) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; DeFrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. Gaussian, Inc., Pittsburg, PA, 1992.