

Calculating the local solvent chemical potential in crystal hydrates

Haluk Resat^{1,*} and Mihaly Mezei²

¹College of Arts and Sciences, Koç University, Istinye, Istanbul, Turkey

²Department of Physiology and Biophysics, Mount Sinai School of Medicine,
New York University, New York, New York 10029-6574

(Received 14 March 2000; revised manuscript received 12 July 2000)

Determining solvation patterns in biological systems is crucial in investigating the functional role water may play in structural stabilization and molecular recognition. Determining whether a particular position would be occupied by a solvent molecule requires the local thermodynamics to be known. In this work we introduce a simple and inexpensive approach based on grand canonical molecular simulations to determine the occupancy factors of the cavities. The method is applied to the test case of the sodium salt of hyaluronic acid. The results agree very well with experimental results and demonstrate the success of the method.

PACS number(s): 87.10.+e, 87.15.He, 87.15.By, 36.20.Ey

I. OVERVIEW

It has been well recognized that water molecules, which strongly interact with proteins and/or nucleic acids, can have functional roles such as structural stabilization, mediating electron transfer and binding of ligands, facilitating biomolecular recognition and catalysis, and aiding complex formation [1]. Unfortunately, mainly due to the relatively higher mobility of waters, the detection of the water locations in biological crystals using conventional x-ray diffraction and NMR approaches is problematic and is often of limited use. Such problems also arise in determining small and mobile molecules in porous solid structures such as zeolites. For these reasons, computational approaches can serve to complement experiments.

One intrinsic difficulty of theoretical approaches in determining the location of small molecules in crystals of biological systems or of porous materials is the irregular shapes and inhomogeneous inner walls of the cavities containing the small molecules. In some cases, the cavities are also disjointed and do not allow for the small molecules to move in between the cavities. In an earlier study, we have called such effects enclosed cavity effects [2]. A biological crystal is an inhomogeneous system, and therefore, the biomolecule (solute molecule) acts as an external source to the solvent molecules and introduces an inhomogeneous field in the cavities. Because of the existence of such an “external” field, it can be expected that dynamic and thermal properties of hydrating water (or other small molecules) in confined spaces could be quite different from that of the bulk system. This makes the *a priori* prediction of how many waters reside in a given cavity and what would be their locations very difficult. Therefore, some “unconventional” approaches are needed to tackle such issues.

We have recently demonstrated that an approach based on grand canonical molecular simulations can be a powerful tool to determine the location of waters in biological systems

[2–5]. The advantages of grand canonical simulations over the more conventional canonical ones can be briefly summarized as [2]: (a) the ability to distribute molecules between cavities without any bias, and (b) the possibility of molecule transfer between the cavities during the simulation run. These differences remove the bias due to the initial configuration and considerably improve the statistical sampling. The only required property that needs to be known is the density of the system, which is used to adjust the chemical potential. In this work, we introduce an approach to compute the local thermodynamic properties of water molecules in crystal hydrates. In this approach the chemical potentials of crystal waters in a certain local region are computed making the density information unnecessary. The computed local thermodynamic properties can then be used to accurately determine the local solvent density in specific regions (cavities) of the crystals. In the following sections we first introduce our approach and then apply it to a sample oligosaccharide crystal hydrate, the sodium salt of hyaluronic acid.

II. GRAND CANONICAL MONTE CARLO SIMULATIONS

In statistical mechanical treatments, such as molecular simulations, the aim is to calculate the ensemble averages $\langle \mathcal{O} \rangle$ of the desired observables $\{\mathcal{O}\}$, and the choice of the extensive variables defines the ensemble used [6]. A grand canonical ensemble simulation is in essence equivalent to a set of appropriately weighted canonical ensemble simulations [3], and the grand canonical partition function for a one-component system $\Xi(\mu, V, T)$ can be expressed in terms of the canonical ensemble partition function $Q(N, V, T)$ as

$$\Xi = \sum_N \frac{z^N}{N!} Q_N, \quad (1)$$

where z is the fugacity (activity) function. Due to this similarity, the grand canonical ensemble simulation methods were developed by generalizing the existing canonical simulation methods [2,7,8] enhanced by cavity-biased insertions [9]. Generalization of the above expression to multicomponents is straightforward. For a two component system, Eq. (1) becomes

*Corresponding author. Present address: Pacific Northwest National Laboratory, P.O. Box 999, Mail stop: K1-83, Richland, WA 99352. Email address: haluk.resat@pnl.gov

$$\Xi(\mu, V, T) = \sum_{\{N_1, N_2\}} \frac{z_1^{N_1} z_2^{N_2}}{N_1! N_2!} Q_N, \quad (2)$$

where $z_i = e^{\beta\mu_i/\Lambda_i^3}$ is the fugacity function of the i th component with the chemical potential μ_i and thermal de Broglie wavelength $\Lambda_i = h/\sqrt{2\pi m_i kT}$. In this case the ensemble average of an observable $\mathcal{O}$ can be expressed as

$$\overline{\mathcal{O}} = \frac{1}{\Xi} \sum_{\{N_1, N_2\}} \frac{z_1^{N_1} z_2^{N_2}}{N_1! N_2!} \overline{\mathcal{O}}_N, \quad (3)$$

where N is the total number of molecules, and $\Sigma\{N_1, N_2\}$ denotes an unrestricted sum over N_1 and N_2 .

In this study we use grand canonical Monte Carlo (GCMC) simulations and we follow Adams's formulation [3,7]. If one defines a “ B ” parameter for the i th component as $B_i = \beta\mu_i + \ln(V/\Lambda_i^3)$, the grand canonical partition function for a two-component system can be written as

$$\Xi = \sum_{\{N_1, N_2\}} \frac{e^{N_1 B_1 + N_2 B_2}}{N_1! N_2!} \int d\tau_N e^{-\beta U_N}, \quad (4)$$

where $d\tau_N = d\Gamma_N/V^N$ is the normalized volume element (i.e., when integrated it gives 1). Therefore, in contrast to the Boltzmann factor $e^{-\beta U_N}$ used in the canonical ensemble, in the grand canonical simulations one utilizes the probability factor $e^{NB - \beta U}/N!$ in generating the Markov chain. In a practical implementation, one tunes the B parameters, i.e., effectively the chemical potentials, until the desired composition is obtained.

In the “traditional” approach described above, an overall solvent density needs to be defined to adjust the chemical potential of solvent molecules. How many solvent molecules a specific cavity contains is found by running a grand canonical molecular simulation using this chemical potential and counting the number of solvent molecules in the investigated region. It should be noted that the cavities observed in biomolecular systems can be quite small; they can contain as few as a single water molecule. Because of the inhomogeneity due to the solute's field, guessing the number of solvent molecules using the cavity volume as the only criteria may give seriously wrong results [10]. The question of how many solvent molecules occupy a given cavity on average can only be answered by determining the local thermodynamic properties of the system. This requires the information about the free energy cost of including or removing water molecule(s) from a specified region. In the following section, we introduce a simple approach which allows for such detailed local population analysis.

A. Local solvent chemical potential

Since the free-energy cost of adding an additional water molecule to a cavity would depend on how many waters it already contains, the local chemical potential of the waters in a specific cavity is a function of the local solvent density. If the thermodynamic properties of occupation of the cavity by waters is known, one can compare the density-dependent local chemical potential with that of a reference system, the *bath* or the *reservoir* (which in our case would be the solu-

tion in which the crystal was grown), and decide on the optimal number of waters for the cavity. One way of obtaining the local free energies is by calculating the incremental free energy of adding an additional solvent molecule into the cavity [10]. This, in principle, can be done by performing a series of free-energy simulations using conventional techniques such as free-energy perturbation or thermodynamic integration methods. However, such free-energy simulations are computationally expensive and have to be performed separately for each cavity. Also, they provide results which are meaningful only for full occupancies. The approach developed in this paper allows for the calculation of the necessary free-energy information in a series of much less expensive grand canonical simulations. Our approach can provide the results simultaneously for each cavity present in the system or one can employ computational tricks such as the color method introduced below and emphasize the sampling in a small region. Such tricks improve the sampling efficiency and reduce the required computational expense. Furthermore, the grand canonical simulations would provide the chemical potential as a continuous function of the density, i.e., statistical fractional occupancies can be also dealt with.

B. Color-bias sampling

To compute the properties of solvent molecules in a specific cavity, one has to find ways to label those molecules as distinct from other solvent molecules. This can be achieved by attaching a *color* label to the waters. The fictitious color label does not change the system Hamiltonian but provides a means of defining to which cavity a certain water molecule belongs. It is essentially a tool where tags are attached to waters allowing the treatment of the system as a multicomponent solvent system. Color labels are used to keep the waters in the specified region from mixing with the remaining water molecules. Tags also specify if a water is deletable and/or insertable. Therefore, the color labeling effectively transforms a one solvent component system into a two-component (or multicomponent) one. In this paper we label the waters in the designated region (call it region C) with color label R , and the waters in the other parts of the system will carry the color label O . All waters with label R reside in region C at the beginning of the simulation, and the same cavity contains no O -labeled water. All R -labeled waters can be inserted and/or deleted during the simulation. With a proper choice of the bias potential (see below), insertion or deletion of the O -labeled waters can be disallowed. Since the molecules can get transferred from one cavity to another during the simulation, obtaining correct statistical data requires the implementation of the necessary procedures to keep R -labeled water in region C and not to allow O -labeled water to enter the cavity. This can be achieved using a bias (umbrella) potential. In biased sampling [11], one writes the interaction potential as $U = U + U_b - U_b \equiv U' - U_b$ where U_b is the bias potential and samples the system using the modified potential U' . The ensemble average of a quantity $\langle \mathcal{O} \rangle$ then can be calculated using

$$\langle \mathcal{O} \rangle = \frac{\langle \mathcal{O} e^{\beta U_b} \rangle'}{\langle e^{\beta U_b} \rangle'}. \quad (5)$$

In the above equation, the symbol $\langle \dots \rangle'$ denotes an en-

semble average using the modified potential U' . The bias potential can, in principle, be defined in any way but in this paper we choose a simple bias potential which is zero inside the desired region and has a constant value outside it,

$$e^{\beta U_b} = \begin{cases} 1 & \text{if } R\text{-labeled water is in } C \text{ region,} \\ \gamma_1 & \text{if } R\text{-labeled water is outside } C \text{ region,} \\ \gamma_2 & \text{if } O\text{-labeled water is in } C \text{ region,} \\ 1 & \text{if } O\text{-labeled water is outside } C \text{ region.} \end{cases} \quad (6)$$

If γ values are chosen to be very large (as done in this study), $\gamma_1, \gamma_2 \rightarrow \infty$, a hard-wall potential is obtained which ensures that the molecules would be kept in their assigned regions. It can be noted that the above set is only a particular choice, soft-core potentials or multiregional potentials can also be used [12,13].

III. COMPUTATIONS

The above outlined approach was applied to determine the thermodynamic properties of waters in orthorhombic crystals of the sodium salt of hyaluronic acid (HYA), a system that we have investigated before [2]. HYA is a linear polydisaccharide of the form $(\text{-GCU-NAG-})_n$, where GCU is glucuronic acid and NAG is *N*-acetylglucosamine [14]. Each unit cell of HYA, with dimensions 11.53, 9.89, and 33.86 Å, consists of four asymmetric subunits. In the unit cell, there are eight residues each of GCU and NAG, eight sodiums, and 16 experimentally determined waters. The linear polydisaccharide chains form two antiparallel left-handed helices. Because of the crystal symmetry, only two of the sodiums and four of the waters are unique (i.e., each subunit contains two Na^+ and four H_2O). To obtain a rectangular simulation cell with less size asymmetry, the crystal unit cell was replicated in the x and y directions to form a bigger simulation unit cell with dimensions 23.06, 19.78, and 33.86 Å. Thus our simulation unit cell contained 16 asymmetric subunits. One of these subunits was assigned as region- C and contained the R -labeled water. Similarly, the remaining subunits were labeled as the region containing the O -labeled water.

As has already been noted before [13], the hard-wall bias potential given in Eq. (6) might introduce errors when the water distribution function is continuous across the boundary of region C . An example of such a case would be if a cavity with assigned borders has surface openings through which waters may leave and enter the region. Since the local thermodynamic properties can strongly depend on the location, the results may be effected by the chosen boundary. However, a close inspection of HYA crystal structure reveals that the cavities containing waters are quite isolated from each other. Therefore, in our case defining an isolated region- C and keeping the labeled waters in that space using a hard-wall bias potential should not have any important effect on the computed chemical potentials. To make sure that this is the case, the waters were monitored during simulations and attempts to cross the boundary were rejected. Such instances occurred only a handful of times (out of millions of trials) when the number of waters in region C was much higher than the available space can realistically accommodate.

The GCMC simulations were performed starting with the

experimental structure. The oligosaccharide polymer and sodium ions were kept fixed at their experimental positions. The Markov chain was generated by alternating between the regular displacement moves and the insertion and/or deletion attempts for water molecules. Color-biased sampling discussed earlier made it possible to treat the system in such a way that the number of water molecules was allowed to fluctuate only in the R -region. Even though the waters in the O -labeled region had the full movement freedom, the bias potential constrains the system in such a way that the total number of O -labeled waters is kept constant. In this aspect, our simulation method is a mixture of grand canonical and canonical ensembles. The temperature was 300 K. The cavity-biased insertions were utilized [9]. To define the cavities, the van der Waals surface of the polymer chains and of the sodiums was created using a solvent probe diameter of 2.4 Å. Employed interaction parameters, charges, and the other details of the simulations can be found in Ref. [2]. A series of GCMC simulations were performed for different values of the B parameter (i.e., the chemical potential). At each B value, the system was equilibrated by running a simulation and observing the equilibration characteristics. Once the system was found to be equilibrated, a run to collect data was started. Data collection runs were performed for 15 to 20 million Monte Carlo (MC) steps. In all cases, the runs were well converged and the approximate error bar in the determined $\bar{N}$, the average number of water molecules, is less than 0.1.

It is necessary to use the chemical potential of waters in experimental conditions as the reference value. We can, however, approximate this value with that of bulk water. For this purpose, the chemical potential of bulk water at $T = 300$ K and at a density of 1 g/cm^3 was also calculated using the GCMC simulation, and it was found to be -10.9 kcal/mol . Note this is the absolute chemical potential corresponding to an excess (over the ideal gas value) chemical potential of -6.3 kcal/mol , in good agreement with earlier studies [16].

IV. RESULTS AND DISCUSSION

Figure 1 shows the chemical potential of waters in one of the subunits of the HYA crystal as a function of number of waters in that subunit. The constant line is the chemical potential of the bulk water. As the figure shows, the local water chemical potential is a monotonously increasing function of the density. In the x-ray experiments, it was found that each subunit contains four waters [14]. We observe that the free energy of setting three waters into the subunit of the HYA crystal hydrate is very favorable and that there is a large increase in the local chemical potential when the number of waters in the subunit is slightly larger than 4. The average number of waters in one subunit can be estimated as $\bar{N} = 4.1$ by equating the chemical potential of the waters in the cavity to that of the bulk water. Therefore, it is expected that most of the subunits will have four waters and some subunits may have more, most likely five waters. This finding is actually consistent with our previous study. In Ref. [2] it was observed that, even though it may have a low occupancy factor, there could be a fifth water in each subunit. In the analysis of the x-ray diffraction spectra, the sodium ions

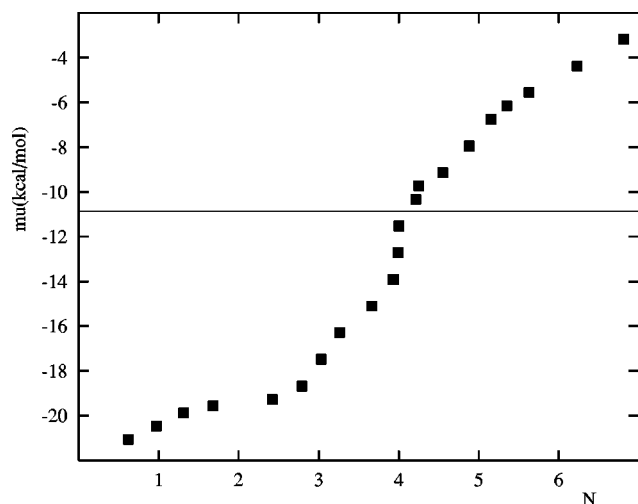


FIG. 1. Chemical potential of waters in a subunit in kcal/mol. N is the number of water molecules in a subunit.

were modeled as having a coordination of six. This built in condition may not have allowed detection of the fifth water location. It has actually been observed in ion chromatography experiments of polymers that the sodium ion can be coordinated to seven oxygens [15].

Note that the function plotted in Fig. 1 shows the average chemical potential as a function of number of waters in the subunit. Therefore, if μ_{k-1} represents the chemical potential when there are $(k-1)$ waters in the subunit, then the free energy cost of adding the k th water into the subunit ΔF_k would be $k(\mu_k - \mu_{k-1}) + \mu_{k-1}$. Table I reports the estimated free-energy costs of adding additional waters into the subunit. The required incremental free energy of adding the fourth water is positive but adding this water is still thermodynamically favorable. The addition of the fifth water further increases the free energy in such a way that the chemical potential of the waters in the cavity becomes higher than the chemical potential of the bulk water. Therefore, the addition of the fifth water is not favored. Thus, thermally accessible configurations with five waters are most likely short-lived intermediate states of the system.

To investigate the effect of keeping the sodium ions fixed during the simulations, we ran a second series of GCMC simulations allowing the sodium ions to move within the

TABLE I. Incremental free-energy cost of adding solvent molecules into the cavity, and the density-dependent solvent chemical potential. μ_N is the water chemical potential if there are N waters in the subunit. μ_N values were extrapolated from the data reported in Fig. 1. ΔF_N is the free-energy cost of adding the N th water when there are already $(N-1)$ waters in the subunit. μ_N and ΔF_N are in kcal/mol.

N	μ_N	ΔF_N
1	-20.4	-20.4
2	-19.4	-18.4
3	-17.6	-14.0
4	-11.5	6.8
5	-7.4	9.0
6	-4.8	8.2

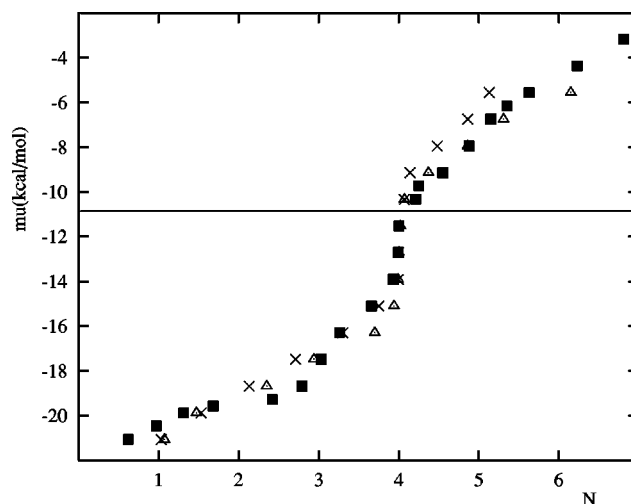


FIG. 2. Chemical potential of waters in a subunit. Comparison of the various simulation approaches. Rectangles: Fixed sodiums with a solvent probe radius of 1.2 Å. Triangles: Mobile sodiums with a solvent probe radius of 1.2 Å. Crosses: Mobile sodiums with a solvent probe radius of 0.6 Å.

cavities they reside. The results were in good agreement with the first case, Fig. 2. We have also investigated if the way the cavities are defined would have an impact on the free-energy results. To this effect, the simulations with mobile sodium ions were repeated. However, this time the van der Waals surface of the polymer chains and of the sodiums was created using a solvent probe diameter of 1.2 Å. The results for this test case study are also reported in Fig. 2. As it can be seen, the results of the three sets of simulations (with or without sodium ion mobility, or with using different solvent probe radius values to define the molecular surfaces) agree very well, see Fig. 2.

We have also investigated the volume available to the solvent molecules by constructing the molecular surfaces. The volumes of the cavities in the crystal depend on how the molecular surface is defined. A probe radius of $R_p = 0.8$ Å seems to be the appropriate value to define the molecular surfaces of biomolecules [5]. With this probe radius, the volume available to solvent molecules in each subunit of the HYA crystal is 116 Å^3 which is roughly the volume of four waters in bulk phase. This observation, too, supports the result that there are four waters in each unit cell. However, calculations based on molecular surfaces to determine the amount of space available to the solvent is very sensitive to the value of the probe radius. For example, if the radius of the probe is increased to 1.2 Å, the volume available to the solvent becomes 21 Å^3 , not enough to accommodate even a single water molecule. Similarly, a smaller probe radius, $R_p = 0.4$ Å, results in a volume of 310 Å^3 which is sufficient to hold ten waters. Such strong sensitivity to the probe radius further confirms the fact that the number of molecules in confined spaces cannot be reliably predicted by using the molecular volumes alone.

In this paper we introduced an approach to calculate the local chemical potential of small molecules in confined spaces. The approach was applied to study the hydration pattern of HYA crystal. As our results show, our theoretical results agree very well with the experimental results. Our

results also demonstrate by example that the method developed here can be a powerful tool by supplementing experimental studies with atomic level structural and thermodynamic data. Our approach is a simple, easy to implement, and computationally inexpensive tool to study the dynamic

and thermodynamic properties of small molecules in confined spaces. A full description of the approach and in-depth analysis of the reported simulations, including the occupation probabilities of solvent residence sites, will be reported elsewhere [17].

-
- [1] E. Westhof, *Water and Biological Macromolecules* (CRC Press, Boca Raton, 1993).
- [2] H. Resat and M. Mezei, *J. Am. Chem. Soc.* **116**, 7451 (1994).
- [3] H. Resat and M. Mezei, *Biophys. J.* **71**, 1179 (1996).
- [4] F. Guarnieri and M. Mezei, *J. Am. Chem. Soc.* **118**, 8493 (1996).
- [5] H. Resat, T. J. Marrone, and J. A. McCammon, *Biophys. J.* **72**, 522 (1997).
- [6] H. L. Friedman, *A Course in Statistical Mechanics* (Prentice-Hall, Englewood Cliffs, NJ, 1985).
- [7] D. J. Adams, *Mol. Phys.* **28**, 1241 (1974); **29**, 307 (1975).
- [8] T. Cagin and B. M. Pettitt, *Mol. Simul.* **6**, 5 (1991).
- [9] M. Mezei, *Mol. Phys.* **61**, 565 (1987); *ibid.* **67**, 1207(E) (1989).
- [10] V. Helms and R. C. Wade, *Proteins: Struct., Funct., Genet.* **32**, 381 (1998).
- [11] G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977).
- [12] J. Hermans and S. Shankar, *Isr. J. Chem.* **27**, 225 (1986); L. Zhang and J. Hermans, *Proteins: Struct., Funct., Genet.* **24**, 433 (1996).
- [13] V. Helms and R. C. Wade, *Biophys. J.* **69**, 810 (1995).
- [14] J. M. Guss, D. W. L. Hukins, P. J. C. Smith, W. T. Winter, S. Arnott, R. Moorhouse, and D. A. Rees, *J. Mol. Biol.* **95**, 359 (1975).
- [15] G. von Helden, T. Wyttenbach, and M. T. Bowers, *Science* **267**, 1483 (1995).
- [16] W. L. Jorgensen, J. F. Blake, and J. K. Buckner, *Chem. Phys.* **129**, 193 (1989); J. Quintana and A. D. J. Haymet, *Chem. Phys. Lett.* **189**, 273 (1992).
- [17] H. Resat and M. Mezei (unpublished).