

Mini-grand canonical ensemble: Chemical potential in the solvation shell

Purushottam D. Dixit,^{1,a)} Artee Bansal,² Walter G. Chapman,² and Dilip Asthagiri^{2,a)}

¹Department of Systems Biology, Columbia University, New York City, New York 10032, USA

²Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77251, USA

(Received 27 June 2017; accepted 4 October 2017; published online 30 October 2017)

Quantifying the statistics of occupancy of solvent molecules in the vicinity of solutes is central to our understanding of solvation phenomena. Number fluctuations in small solvation shells around solutes cannot be described within the *macroscopic* grand canonical framework using a single chemical potential that represents the solvent bath. In this communication, we hypothesize that molecular-sized observation volumes such as solvation shells are best described by coupling the solvation shell with a mixture of particle baths each with its own chemical potential. We confirm our hypotheses by studying the enhanced fluctuations in the occupancy statistics of hard sphere solvent particles around a distinguished hard sphere solute particle. Connections with established theories of solvation are also discussed. *Published by AIP Publishing.* <https://doi.org/10.1063/1.4993178>

A quantitative description of thermodynamics at the molecular scale is of crucial importance in many biological as well as nano-technological systems.¹ While traditional statistical mechanical ensembles can describe macroscopic systems, they are inadequate in describing “small systems,” that is, systems where the range of inter-particle interactions are comparable to the extent of the system itself and thus the system cannot couple in a weak fashion with its surrounding. For example, we recently showed that the probability distribution $p(\bar{r})$ of degrees of freedom \bar{r} of a small system exchanging energy with a surrounding bath is *not solely* determined by its Hamiltonian $H_{\text{sys}}(r)$ and a unique temperature that describes system-bath interactions.^{2,3} A key observation in those studies was that in the absence of weak coupling, the degrees of freedom of a small system that is exchanging energy with its surrounding can be accurately described by allowing its temperature to fluctuate. Mathematically,

$$p(\bar{r}) = \int p(r|\beta) \cdot p(\beta) d\beta, \quad (1)$$

where $p(r|\beta)$ is the usual Gibbs-Boltzmann canonical ensemble distribution. Note that for weak coupling, as is usually assumed in macroscopic thermodynamics, $p(\beta)$ would be a delta function, with β being the same as the bath temperature. The functional form of the distribution $p(\beta)$ over bath temperatures was estimated using maximum entropy arguments.^{2,3}

The inability of small systems to couple weakly with their surroundings is likely to hold true for all statistical mechanical ensembles. In other words, if the system-bath exchanges are comparable to the corresponding property of the system, statistical mechanics based on average extensive quantities is expected to fail. In this work, we study a grand-canonical small system. We study the statistics of number fluctuations in

solvation shells of solute molecules. Understanding the number statistics at the molecular scale is of particular interest in biochemistry; typically small molecules bind to biological macromolecules such as proteins and nucleic acid polymers (RNA and DNA) in binding sites comparable to the size of the small molecule and whose chemical composition can fluctuate. Indeed, the competitive binding of ligands to such sites directly depends on the statistics of ligand occupancy in the binding site.⁴⁻⁷

We work with a hard sphere system to avoid confounding effects due to energetic interactions. For concreteness, we consider a bath of $N \gg 1$ hard sphere particles with one solute particle fixed at the origin. Let the radius of each hard sphere particle, including the solute, be r_p . Within the bath, imagine a solvation shell of radius R around the solute. The number n of solvent particles inside the shell fluctuates as the bath samples configurations according to the microcanonical ensemble (see Fig. 1). The probability $p(n|R)$ of observing n solvent particles in the solvation shell is a centrally important quantity in the study of solvation phenomena, such as ion hydration^{5,8} and the hydrophobic effect.⁹

We note that if $R \gg r_p$, the grand canonical ensemble predicts that the probability of observing n particles in the solvation shell is

$$p_\mu(n|R) \sim e^{-F(n)+\mu n}, \quad (2)$$

where $F(n)$ is the free energy of assembling n solvent particles around the solute in a shell of radius R *in the absence of the rest of the solvent*. Without the loss of generality, we have assumed that $\beta = 1$. Here, μ is the chemical potential that dictates the coupling between the solvation shell and the bath. Note that for the grand canonical ensemble, $F(n)$ only depends on the configurations of the solvation shell and does not depend on the nature of exchange of particles between the solvation shell and the bath and on the chemical potential of the bath.

A key feature of the grand canonical description is that a single bath parameter μ describes all moments of the

^{a)}Authors to whom correspondence should be addressed: dixitpd@gmail.com and dna6@rice.edu

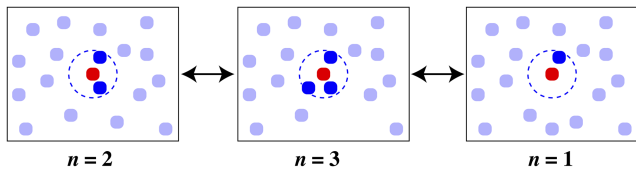


FIG. 1. Illustration of the number fluctuations in a solvation shell (dashed circle) of a solute (red circle) as the solvation shell samples configurations from the microcanonical ensemble.

number distribution, $p_{\mu}(n|R)$, as derivatives of the grand canonical partition function.¹

Is the grand canonical prescription accurate when the size of the solvation shell is comparable to the size of the particle ($R \sim r_p$)? In Fig. 2, we show the distribution $p(n|R)$ of observing n hard sphere solvent particles in a solvation shell of radius $R = 2.2 \times r_p$ around a distinguished hard sphere solute particle of the same size (black circles). The reduced density of the solvent is $8\rho r_p^3 = 0.9$. The average number of solvent particles in the solvation shell is $\langle n \rangle \approx 4.5$. Note that in order to ensure efficient sampling of rarely occupied states, we employ an ensemble reweighing technique that has been employed earlier to study solvent occupancy statistics in solvation shells as diverse as water,¹⁰ hydrated ions,¹¹ and hard-sphere mixtures with large solute-solvent size asymmetry.^{12,13}

Next, in order to find the best grand-canonical description of the solvation shell, we conducted grand canonical Monte Carlo simulations in a solvation shell of radius $R = 2.2 \times r_p$ for $\mu \in [-10, 5]$ with an interval of $\delta\mu = 0.05$ (see inset for $\langle n \rangle$ and $\sqrt{\langle n^2 \rangle - \langle n \rangle^2}$ as a function of μ). Then, we chose the value of $\mu^* \approx -0.6$ that reproduced the average number of particles $\langle n \rangle \approx 4.5$ in the solvation shell. The dashed blue line shows the grand canonical estimate of the number distribution

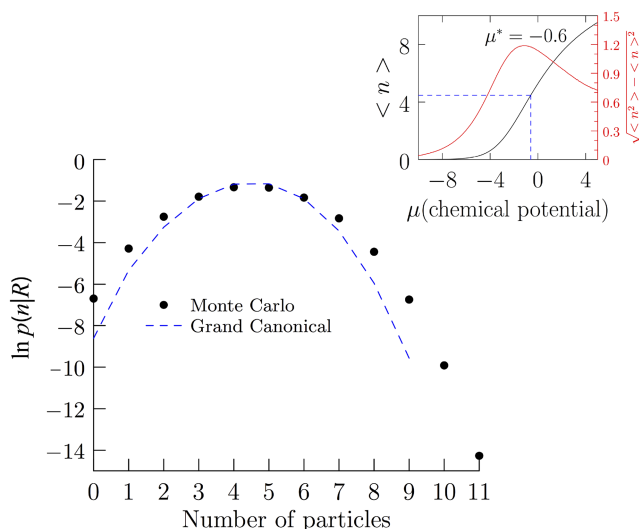


FIG. 2. Distribution $p(n|R)$ of the number n of solvent particles inside the solvation shell of radius $R = 2.2 \times r_p$ of a spherical solute particle with radius r_p in a solution of three dimensional hard sphere solvent particles of radius r (black circles). The reduced density of the solvent is $\rho\sigma^3 = 0.9$. Dashed blue line shows the distribution $p_{\mu=\mu^*}(n|R)$ of a grand canonical ensemble simulation where the chemical potential $\mu^* \approx -0.6$ is adjusted to reproduce the average occupancy $\langle n \rangle \approx 4.5$. The inset shows the average occupancy $\langle n \rangle$ and the standard deviation in occupancy $\sqrt{\langle n^2 \rangle - \langle n \rangle^2}$ as a function of the chemical potential μ .

$p_{\mu=\mu^*}(n|R)$. Notably, the grand canonical ensemble predicts a distribution with a lower variance $\langle n^2 \rangle - \langle n \rangle^2$ in the occupancy statistics compared to simulation results.

What factors lead to these enhanced number fluctuations in the solvation shell compared to the grand-canonical ensemble? For molecular-sized solvation shells, local density fluctuations in the solvent are of the same magnitude as the occupancy of the solvation shell itself. Moreover, density fluctuations just outside the solvation shell will depend on the occupancy of the solvation shell. This implies that the work required to transfer a solvent particle across the boundary of the solvation shell will depend on both the local density of solvent particles just outside the solvation shell as well as the density of solvent particles in the solvation shell.^{5,8,13} As a result, a single chemical potential cannot represent the exchange of solvent particles between the solvation shell (“system”) and the rest of the solvent (“bath”). Notably, the grand canonical picture does not allow us to separately tune the mean occupancy as well as the standard deviation around the mean.

How do we capture these enhanced number fluctuations in the solvation shell? Based on our previous work with the canonical ensemble,^{2,3} we propose a generalization of the grand canonical picture. We hypothesize that a small solvation shell that exchanges particles with a surrounding bath can be represented as a solvation shell that is in contact with multiple baths, each characterized by a unique chemical potential μ . Let $P(\mu)$ be the probability distribution over the baths. We can obtain the occupancy distribution of solvent particles by marginalizing over bath chemical potentials

$$p_{ss}(n|R) = \int P(\mu) \times p_{\mu}(n|R) d\mu, \quad (3)$$

where $p_{\mu}(n|R)$ is given by Eq. (2).

What is the functional form of $P(\mu)$? One numerical approach, inspired by research in image processing,¹⁴ is to constrain the L_2 error between the observed distribution $p(n|R)$ and the predicted distribution $p_{ss}(n|R)$ while maximizing the entropy of $P(\mu)$. While this approach should lead to accurate predictions of $p(n|R)$,¹⁴ the numerically inferred distribution $P(\mu)$ offers little physical clarity. Another, more conceptual approach is to motivate the functional form of $P(\mu)$ using first principles. Previously, we have shown using maximum entropy arguments² that the distribution of inverse temperatures $P(\beta)$ of a small system can be described as an inverse gamma distribution. But this functional form is not suitable for $P(\mu)$, because while β is always positive, μ can take both positive and negative values. The inverse gamma distribution does not support negative arguments. However, the activity $z = \exp(\mu)$ only takes positive values. Thus, as a first guess, we assume that the inverse gamma distribution describes the distribution of bath activities $z = \exp(\mu)$. Specifically, we assume that

$$P(\mu) = \frac{e^{(-\lambda_1 e^{-\mu} - \lambda_2 \mu)}}{\Gamma(\lambda_2) \lambda_1^{-\lambda_2}}. \quad (4)$$

As we see in Fig. 3, this particular functional form accurately describes the solvent number fluctuations around a solute molecule. We leave it to future studies to better explore the

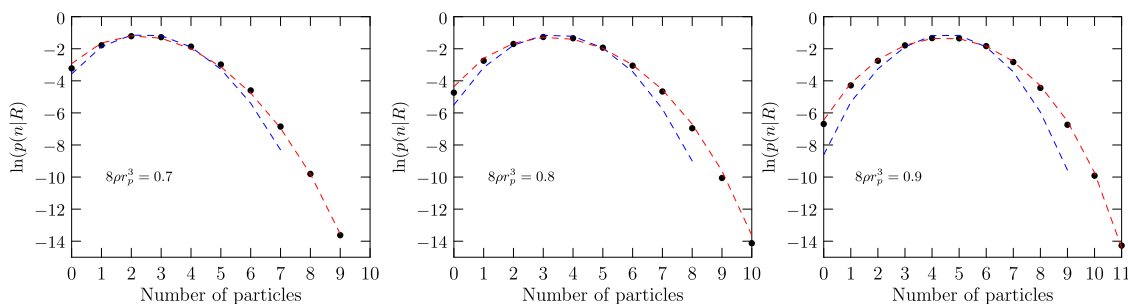


FIG. 3. The probability $p(n|R)$ of observing n solvent particles in a solvation shell of radius $R = 2.2 \times r_p$ of a solute particle of size r_p at reduced densities $8\rho_p^3 = 0.7, 0.8$, and 0.9 (black circles). The grand-canonical prediction $p_{\mu=\mu^*}(n|R)$ at each density is shown with dashed blue lines. The prediction $p_{ss}(n|R)$ is shown with dashed red lines.

relationship between $P(\mu)$ and the nature of the interactions between the solvation shell and the surrounding bath.

Next, we test whether Eq. (3) can accurately capture the number statistics in the solvation shell. In Fig. 3, we show $p(n|R)$ at reduced densities $8\rho_p^3 = 0.7, 0.8$, and 0.9 , respectively. At each reduced density, we find the chemical potential μ^* of the grand canonical ensemble that matches the mean occupancy $\langle n \rangle$. From Fig. 3, it is clear that for molecular-sized solvation shells, the occupancy statistics predicted by the grand canonical ensemble [Eq. (2)] fails to capture the variation in the distribution of the number of solvent particles in the solvation shell.

Using the probabilities $p_\mu(n|R)$ of observing n solvent particles in the solvation shell in grand canonical simulations, at each reduced density $8\rho_p^3$, and a simulated annealing scheme, we determine the parameters λ_1 and λ_2 [Eq. (4)] that lead to the lowest error between $\log p(n|R)$ and $\log p_{ss}(n|R)$. The dashed red lines show the predicted probability $p_{ss}(n|R)$ using Eq. (3). It is clear that the distribution $p_{ss}(n|R)$ can capture the entire distribution of solvent occupancy numbers very well for multiple solvent densities. We emphasize that

the good fit is not simply a result of including two parameters [Eq. (4)] versus one parameter, the bath chemical potential. Indeed, using a two-moment maximum entropy description with either a Gibbs prior or a flat prior with the mean occupancy and the variance in occupancy as constraints also fails to capture $p(n|R)$ in its entirety.^{15,16} We return to this point below.

We also investigate the approach to the eventual macroscopic grand canonical description. As before, using the ensemble reweighing technique, we estimate the number distribution $p(n|R)$ for solvation shells of size $R = 2.2 \times r_p, 2.4 \times r_p, 2.6 \times r_p$, and $2.8 \times r_p$. The reduced density of the solvent was fixed at $8\rho_p^3 = 0.8$. Next, for each solvation shell of radius R , we perform grand canonical Monte Carlo simulations and estimate the μ^* that reproduced the mean occupancy $\langle n \rangle$. In Fig. 4, we show $p(n|R)$ in black dots and the corresponding grand canonical estimate $p_{\mu=\mu^*}(n|R)$ in dashed blue lines. As expected, we observe that the discrepancies between $p(n|R)$ and the grand canonical estimate $p_{\mu=\mu^*}(n|R)$ decrease as R increases. Notably, as shown in Fig. 5, the distribution of chemical potentials $P(\mu)$ becomes narrower as the radius of the

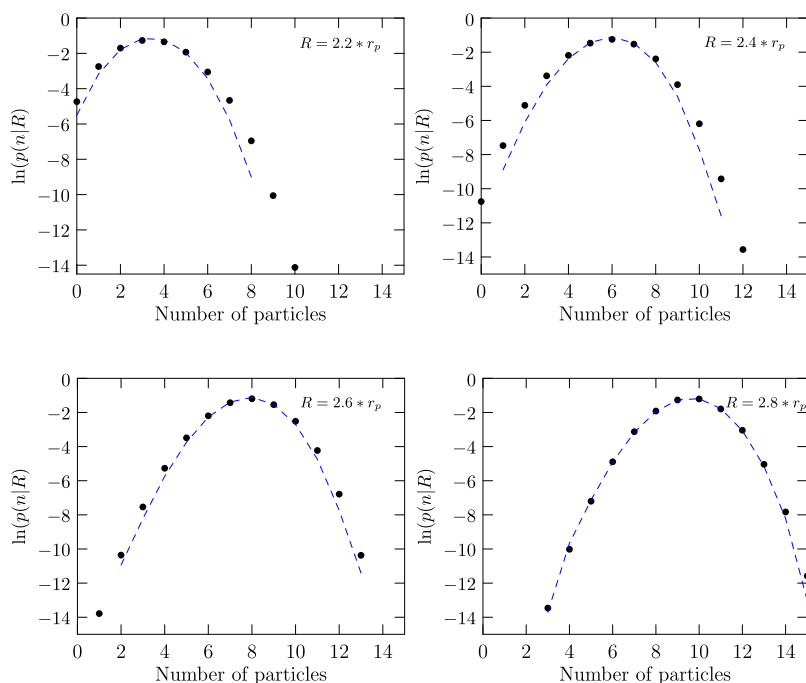


FIG. 4. The probability $p(n|R)$ of observing n solvent particles in a solvation shell of radius $R = 2.2 \times r_p, 2.4 \times r_p, 2.6 \times r_p$, and $2.8 \times r_p$ (black dots) and the corresponding grand canonical Monte Carlo estimate $p_{\mu=\mu^*}(n|R)$ (dashed blue line). The reduced density of the solvent is $8\rho_p^3 = 0.8$ in all 4 cases.

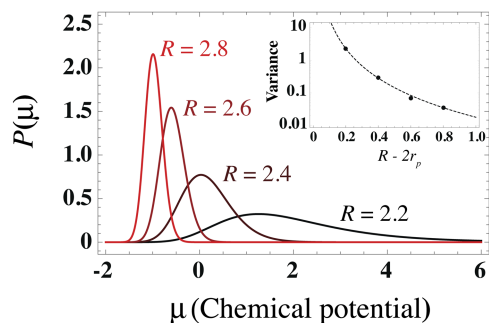


FIG. 5. The distribution of chemical potential $P(\mu)$ estimated for four different observation volumes of increasing radius R used in Fig. 4. The inset shows the variance of $P(\mu)$ as a function of the radius R . Notably, the variance scales as $(R - 2r_p)^{-3}$.

observation volume increases. This makes good physical sense since as the solvation shell approaches the thermodynamic limit, i.e., $R \gg r_p$, $P(\mu)$ should approach a Dirac delta function. Interestingly, the variance in chemical potential $\langle \mu^2 \rangle - \langle \mu \rangle^2$ monotonically approaches zero, the macroscopic limit, with a scaling $\sim (R - 2r_p)^{-3}$. We leave it to future studies to systematically explore the approach to macroscopic behavior for different densities, shapes of observation volumes, and in the presence of attractive interactions.

The statistics of occupancy of the solvation shell of a solute molecule is of central importance in phenomena ranging from hydration^{5,8,11} to understanding small molecule binding to specific sites in proteins and nucleic acids.⁴⁻⁷ While the solvation shell is a system open to the exchange of particles, the tight coupling with the environment does not allow one to use the grand canonical framework with a single chemical potential for both molecules.

Here we have proposed to generalize the grand canonical picture by coupling the solvation shell to multiple solvent baths, each with its own chemical potential. This approach accurately captures the statistics $p(n|R)$ of the number n of solvent molecules in the solvation shell of radius R . As the size of the solvation shell increases compared to the size of the solute particle, the agreement with the grand canonical description is restored. In this work, we approximated the distribution $P(z)$ of the bath chemical activities with an inverse gamma distribution [and correspondingly $P(\mu)$ by Eq. (4)]. However, relevant experimental constraints can inform the choice of $P(\mu)$ as well. For example, the first and the second moments of number fluctuations in the solvation shell can be obtained directly from experimental measurements of the structure factor.¹⁷ $P(\mu)$ can then be obtained by maximizing its entropy while imposing these constraints.

How do we reconcile our development with previous work on understanding solvation phenomena? The non-constant nature of the bath chemical potential is well recognized in the quasi-chemical approach.¹⁸ We describe connections with two previous studies. Earlier, building on the quasicheical approach, Merchant *et al.* presented a *molecular aufbau* approach to build the solvation shell around a solute, an approach that makes explicit the thermodynamic reorganization of the solvation shell upon addition of subsequent solvent molecules.^{8,11} Specifically, Merchant *et al.* showed

that^{8,11}

$$\beta[G(n) - G(n-1)] = \log \frac{p^0(n|R)/p^0(n-1|R)}{p(n|R)/p(n-1|R)}, \quad (5)$$

where $p^0(n|R)$ denotes the probability of observing n solvent particles in the solvation shell in the absence of the solute and $G(n)$ is the excess free energy of the solute in the n th coordinate state.

From Eq. (5), it is clear that the work required to insert a solvent particle in the solvation shell itself depends on the current occupancy of the solvation shell. In contrast, for a macroscopic system, the work in Eq. (5) will be independent of the occupancy of the solvation shell and equal to the chemical potential. Thus, on the one hand, the quasicheical approach provides a detailed microscopic description of how the solvent chemical potential depends on the exact solvation state of the solute. The current approach, on the other hand, gives a global picture of the fluctuation in the chemical potential averaged over all solvation states. We hypothesize that a wide $P(\mu)$ may reflect enhanced density fluctuations of the solvent just outside of the solvation shell, and as the width of $P(\mu)$ decreases, the density fluctuations just outside the solvation shell ought to become indistinguishable from the bulk solvent. In the future, we plan to investigate these phenomena.

An approach more closely related to the current work is that of Bansal *et al.*,^{12,13} who analyzed the solvation shell for hard sphere solutes explicitly in terms of the interactions within the cluster around the solute, interactions of the cluster with the rest of the medium, and the interactions for the bulk solvent molecules which are not in the cluster. Based on the partition function in the canonical ensemble for the system which has n molecules within the spherical shell around the solute, they obtained the probabilities $p(n|R)$ as (in the present notation)

$$p(n|R) \sim p_{\mu_p}(n|R) \times e^{\Omega\sigma_n}, \quad (6)$$

where μ_p is the excess chemical potential of the solvent particle and $p_{\mu_p}(n|R)$ is given by Eq. (2) with the μ equal to the chemical potential of the solvent. $\Omega\sigma_n$ approximately represents the field imposed by the bulk solvent on the solute-solvent cluster. The field was explicitly recognized as a surface interaction term that depends on the occupancy of the surface sites in the cluster around the solute and represented as follows:

$$\Omega\sigma_n = \zeta_1 \cdot n^2 + \zeta_2 \cdot n. \quad (7)$$

In other words, Bansal *et al.*^{12,13} apply a correction to the grand canonical ensemble explicitly for each coordination state (see also the work of Reiss and Merry¹⁹). Interestingly, Eq. (6) can also be derived using maximum entropy arguments with the grand canonical distribution $p_{\mu_p}(n|R)$ as the prior (with the μ equal to the chemical potential of the solvent) with mean occupancy and second moment of occupancy as constraints. [Note that absence of the correction term in the work of Bansal *et al.* is equivalent to having $P(\mu)$ as a Dirac delta function centered around the solvent chemical potential in the current work.] The latter point also shows the essential physics captured by the present approach and that of Bansal *et al.* that is absent in two-moment maximum entropy models with either a Gibbs prior or a flat prior. In the current generalization and the approach by Bansal *et al.*, the physics of clustering albeit in the absence

of the external medium is rigorously captured. In the current work, the medium appears as a hierarchy of chemical potential baths, whereas in the approach of Bansal *et al.*, the medium appears as imposing a field. While these pictures are complementary, we believe that the current maximum entropy-based approach may allow us to integrate multiple types of experimental data and reconstruct solvation shell characteristics of realistic solutes.

Anticipating future studies, we note earlier that we had shown that the behavior of small systems that exchange energy with their surroundings can be captured by allowing them to couple with multiple baths each with its own temperature.^{2,3} For open systems with attractive/repulsive interactions, particle number fluctuations will also involve energy exchange with the surrounding bath. We hypothesize that the behavior of such systems, ones that exchange both energy and particles with their surroundings, can be captured by allowing both the temperature and the chemical potential of the bath to fluctuate. Testing this proposal for problems such as characterizing the solvation of states of biomolecules is left for future studies.

ACKNOWLEDGMENTS

A.B., W.G.C., and D.A. gratefully acknowledge support from the Robert A. Welch Foundation (Grant No. C-1241) to Professor Walter G. Chapman. W.G.C. and D.A. additionally acknowledge support from the Abu Dhabi National Oil Company (ADNOC). Finally, we would like to thank the two

reviewers whose comments have substantially improved this manuscript.

- ¹K. Dill and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience* (Garland Science, 2010).
- ²P. D. Dixit, *J. Chem. Phys.* **138**, 184111 (2013).
- ³P. D. Dixit, *Phys. Chem. Chem. Phys.* **17**, 13000 (2015).
- ⁴P. D. Dixit, S. Merchant, and D. Asthagiri, *Biophys. J.* **96**, 2138 (2009).
- ⁵D. Asthagiri *et al.*, *Chem. Phys. Lett.* **485**, 1 (2010).
- ⁶P. D. Dixit and D. Asthagiri, *Biophys. J.* **100**, 1542 (2011).
- ⁷P. D. Dixit and D. Asthagiri, *J. Gen. Physiol.* **137**, 427 (2011).
- ⁸S. Merchant and D. Asthagiri, *J. Chem. Phys.* **130**, 195102 (2009).
- ⁹G. Hummer, S. Garde, A. E. Garcia, M. E. Paulaitis, and L. R. Pratt, *J. Phys. Chem. B* **102**, 10469 (1998).
- ¹⁰S. Merchant, J. K. Shah, and D. Asthagiri, *J. Chem. Phys.* **134**, 124514 (2011).
- ¹¹S. Merchant, P. Dixit, K. Dean, and D. Asthagiri, *J. Chem. Phys.* **135**, 054505 (2011).
- ¹²A. Bansal, D. Asthagiri, K. R. Cox, and W. G. Chapman, *J. Chem. Phys.* **145**, 074904 (2016).
- ¹³A. Bansal, A. Valiya Parambathu, D. Asthagiri, K. R. Cox, and W. G. Chapman, *J. Chem. Phys.* **146**, 164904 (2017).
- ¹⁴S. Gull and J. Skilling, *Commun., Radar Signal Process., IEE Proc. F* **131**, 646 (1984).
- ¹⁵G. Hummer, S. Garde, A. E. Garcia, A. Pohorille, and L. R. Pratt, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 8951 (1996).
- ¹⁶A. Paliwal, D. Asthagiri, L. Pratt, H. Ashbaugh, and M. Paulaitis, *J. Chem. Phys.* **124**, 224502 (2006).
- ¹⁷D. A. McQuarrie, *Statistical Thermodynamics* (Harper and Row, 1973).
- ¹⁸T. L. Beck, M. E. Paulaitis, and L. R. Pratt, *The Potential Distribution Theorem and Models of Molecular Solutions* (Cambridge University Press, 2006).
- ¹⁹H. Reiss and G. A. Merry, *J. Phys. Chem.* **85**, 3313 (1981).