Transfer Free Energies of Test Proteins Into Crowded Protein Solutions Have Simple Dependence on Crowder Concentration

Valery Nguemaha[1], Sanbo Qin[1] and Huan-Xiang Zhou[1,2]*

[1] Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL, United States,
[2] Department of Chemistry and Department of Physics, University of Illinois at Chicago, Chicago, IL, United States

The effects of macromolecular crowding on the thermodynamic properties of test proteins are determined by the latter's transfer free energies from a dilute solution to a crowded solution. The transfer free energies in turn are determined by effective protein-crowder interactions. When these interactions are modeled at the all-atom level, the transfer free energies may defy simple predictions. Here we investigated the dependence of the transfer free energy ($\Delta\mu$) on crowder concentration. We represented both the test protein and the crowder proteins atomistically, and used a general interaction potential consisting of hard-core repulsion, non-polar attraction, and solvent-screened electrostatic terms. The chemical potential was rigorously calculated by FMAP (Qin and Zhou, 2014), which entails expressing the protein-crowder interaction terms as correlation functions and evaluating them via fast Fourier transform (FFT). To high accuracy, the transfer free energy can be decomposed into an excluded-volume component ($\Delta\mu_{e-v}$), arising from the hard-core repulsion, and a soft-attraction component ($\Delta\mu_{s-a}$), arising from non-polar and electrostatic interactions. The decomposition provides physical insight into crowding effects, in particular why such effects are very modest on protein folding stability. Further decomposition of $\Delta\mu_{s-a}$ into non-polar and electrostatic components does not work, because these two types of interactions are highly correlated in contributing to $\Delta\mu_{s-a}$. We found that $\Delta\mu_{e-v}$ fits well to the generalized fundamental measure theory (Qin and Zhou, 2010), which accounts for atomic details of the test protein but approximates the crowder proteins as spherical particles. Most interestingly, $\Delta\mu_{s-a}$ has a nearly linear dependence on crowder concentration. The latter result can be understood within a perturbed virial expansion of $\Delta\mu$ (in powers of crowder concentration), with $\Delta\mu_{e-v}$ as reference. Whereas the second virial coefficient deviates strongly from that of the reference system, higher virial coefficients are close to their reference counterparts, thus leaving the linear term to make the dominant contribution to $\Delta\mu_{s-a}$.

Keywords: macromolecular crowding, transfer free energy, excluded-volume, soft attraction, crowder concentration

INTRODUCTION

It is now well-recognized that "bystander" macromolecules in cellular milieus may significantly influence the biophysical properties of proteins (Zhou et al., 2008; Zhou, 2013; Gnutt and Ebbinghaus, 2016). Such influences can be detected by many experimental observables, including equilibrium sedimentation gradient (Rivas et al., 1999), protein folding and binding stability (Batra et al., 2009a,b; Miklos et al., 2011, 2013; Phillip et al., 2012; Wang et al., 2012; Sarkar et al., 2013), light scattering intensity (Wu and Minton, 2013), small-angle neutron scattering profile (Goldenberg and Argyle, 2014; Banks et al., 2018), and fluorescence resonance energy transfer (FRET) efficiency (Soranno et al., 2014). Theoretically these influences are determined by the transfer free energies of test proteins from a dilute solution to a solution of the macromolecular crowders (Minton, 1983; Zhou et al., 2008; Qin and Zhou, 2009; McGuffee and Elcock, 2010). Since the transfer free energies in turn are determined by the effective protein-crowder interactions, in principle the experimental data contain information about these intermolecular interactions. However, apart from the work of McGuffee and Elcock at "very significant computational expense" (McGuffee and Elcock, 2010), until recently it was only possible to use relatively crude models of protein-crowder interactions for calculating transfer free energies (Minton, 1981; Qin and Zhou, 2010; Kim and Mittal, 2013) or quantitatively modeling crowding effects (Cheung et al., 2005; Minh et al., 2006), thereby limiting our ability to interpret and fully utilize the experimental data. To mitigate this problem, an <u>F</u>FT-based method for <u>M</u>odeling <u>A</u>tomistic <u>P</u>rotein-crowder interactions, or FMAP, has been developed (Qin and Zhou, 2013, 2014). Most recently FMAP was used to quantitatively interpret FRET efficiency data for disordered proteins in the presence of polyethylene glycol (Soranno et al., 2014), implicating mild attraction between the test proteins and the polymer crowder (Nguemaha et al., 2018). Here we used FMAP to calculate the transfer free energy ($\Delta\mu$) of folded and unfolded test proteins from dilute to crowded protein solutions, paying particular attention to the dependence of $\Delta\mu$ on crowder concentration. Even with an all-atom representation for both the test and the crowder proteins, the dependence of $\Delta\mu$ on crowder concentration was found to follow simple relations. We explore the physical reasons for this simple behavior.

As in our previous study (Qin and Zhou, 2014), we assumed a general, implicit-solvent energy function for protein-crowder interactions, consisting of hard-core steric repulsion, non-polar attraction, and solvent-screened electrostatic terms:

$$U_{\text{int}} = U_{\text{st}} + U_{\text{n−a}} + U_{\text{elec}} \qquad (1)$$

Whenever r_{ij}, the distance of any pair of protein-crowder atoms, is less than the sum of their hard-core radii, $(\sigma_{ii} + \sigma_{jj})/2$, the steric term U_{st} goes to ∞. When the test protein is free of such steric clashes with crowder atoms, U_{st} vanishes and the two soft interaction terms come into play. Specifically, the non-polar

attraction has the form of a Lennard-Jones potential:

$$U_{\text{n−a}} = \sum_{ij} 4\varepsilon_{ij}[(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6] = \sum_{ij}(A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^6)$$
$$(2)$$

where ε_{ij} is the magnitude of the non-polar attraction between the i-j pair of atoms. The solvent-screened electrostatic term has the form of a Debye-Hückel potential:

$$U_{\text{elec}} = \sum_{ij} q_i q_j \exp\left(-r_{ij}/\lambda\right)/\kappa r_{ij} \qquad (3)$$

where q_i are atomic charges, and λ and κ are the Debye screening length and the dielectric constant, respectively, of the crowder solution.

FMAP finds the transfer free energy from an average of the Boltzmann factor of the protein-crowder interaction energy (Qin and Zhou, 2013, 2014)

$$\exp\left(-\Delta\mu/k_BT\right) = < \exp[-U_{\text{int}}(\mathbf{R})/k_BT] >_{\mathbf{R},\Omega,\mathbf{c}} \qquad (4)$$
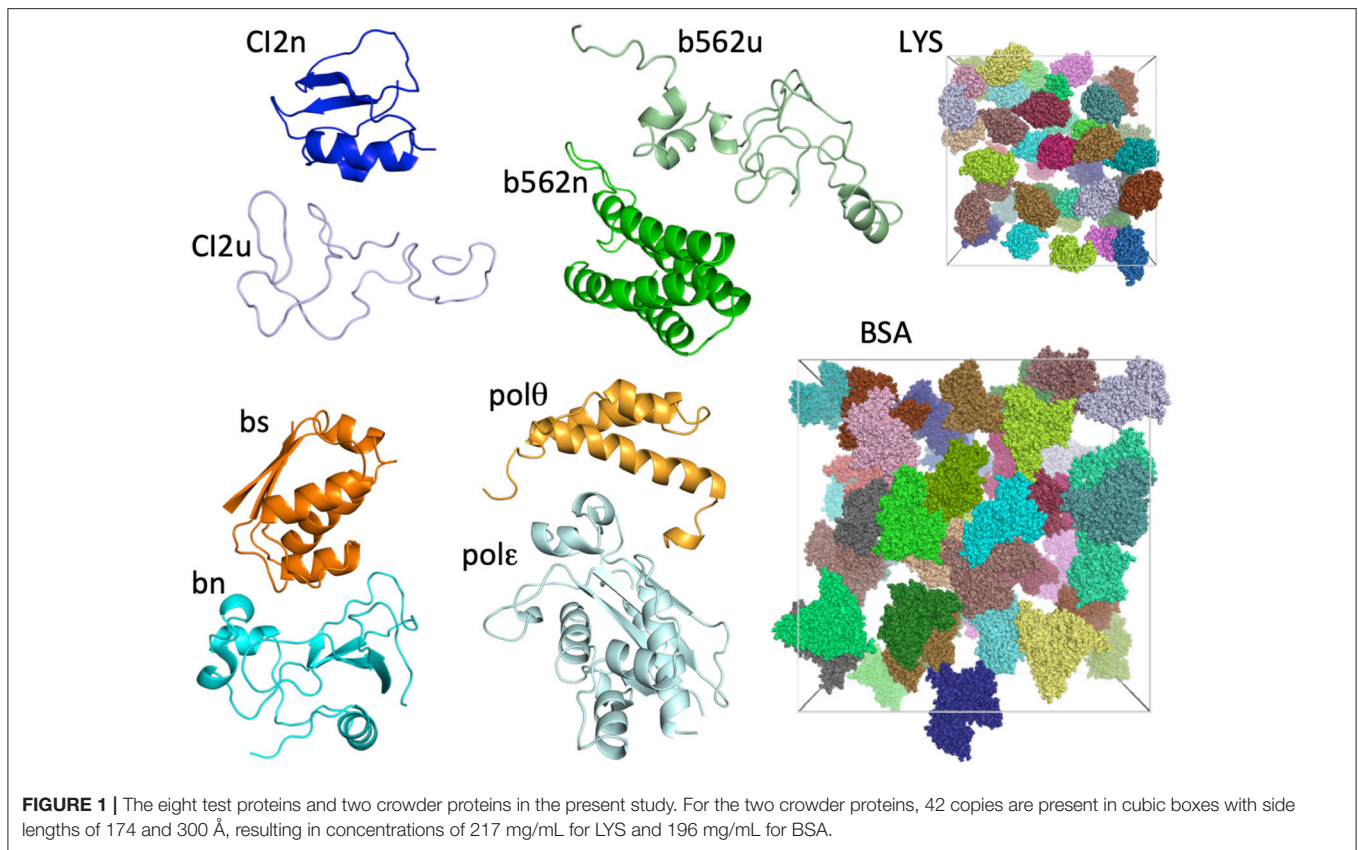
More specifically, the test protein is fictitiously placed into the crowder solution, and the average is taken over the position ("$\mathbf{R}$") and orientation ("Ω") of the test protein and configuration ("$\mathbf{c}$") of the crowders. The average over $\mathbf{R}$ is taken care of by FFT, and is the core component of FMAP. The average over Ω and $\mathbf{c}$ is realized by repeating MFAP calculations over different orientations of the test protein and different configurations of the crowder solution.

Here we obtained the transfer energies of eight test proteins over a wide range of concentrations of two crowder proteins. There are two main findings. First, the transfer free energy can be accurately decomposed into an excluded-volume component ($\Delta\mu_{\text{e−v}}$), arising from the hard-core repulsion, and a soft-attraction component ($\Delta\mu_{\text{s−a}}$), arising from non-polar and electrostatic interactions. Second, whereas the excluded-volume component has a complex dependence on crowder concentration, the soft-attraction component has a nearly linear dependence on crowder concentration. We explain this interesting result by a perturbed virial expansion of $\Delta\mu$.

COMPUTATIONAL METHODS

The eight test proteins studied are: native and unfolded chymotrypsin inhibitor 2 (CI2n and CI2u, respectively), native and unfolded cytochrome b562 (b562n and b562u, respectively), barnase (bn), barstar (bs), and the DNA polymerase III θ and ε subunits (polθ and polε, respectively). As in our previous study (Qin and Zhou, 2014), we represented each protein by a single conformation (**Figure 1**). The two crowder proteins are lysozyme (LYS) and bovine serum albumin (BSA) (**Figure 1**).

To obtain the crowder protein configurations, first hard-sphere simulations were carried out and then the hard spheres in the final snapshots were replaced by randomly oriented protein molecules. The simulations were run using a C++ code (https://cims.nyu.edu/~donev/Packing/C++/), written by Skoge

FIGURE 1 | The eight test proteins and two crowder proteins in the present study. For the two crowder proteins, 42 copies are present in cubic boxes with side lengths of 174 and 300 Å, resulting in concentrations of 217 mg/mL for LYS and 196 mg/mL for BSA.

et al. (Skoge et al., 2006). In short, N spheres in a cubic box were grown from points at a steady rate and underwent ballistic collisions. The box had a side length of 1 and periodic boundary conditions were imposed. The simulations were terminated when the hard spheres grew to a desired radius. Specifically, for the simulations intended for LYS, the final radius was 0.1485, such that the hard-sphere volume fraction at $N = 48$ reached 0.658; for BSA, the final radius was 0.14 and the volume fraction at $N = 48$ was 0.552. Ten replicate simulations were run at each N for replacement into each of the two crowder proteins.

For replacing the hard spheres by protein molecules, the radii of the spheres were scaled to appropriate lengths to allow for the spheres to enclose the proteins. For the simulations intended for LYS, the unit length of the simulation box was scaled to 174 Å, and so the spheres were mapped to a radius of 25.84 Å. For BSA, the corresponding simulation box was scaled to a 300 Å side length, leading to a hard sphere radius of 42.0 Å. These spheres were sufficiently large to enclose the vast majority of the atoms in each crowder protein. The spheres were replaced by protein molecules one at a time. The protein molecules were assigned random orientations, by choosing a random direction for a unit vector attached to the protein and rotating the protein around the unit vector by a random angle between 0 and 360° (Qin et al., 2011). When placing a new protein molecule, random orientations were repeatedly chosen until it did not clash with any of the protein molecules already placed (including their periodic images). The threshold for clash was 4.0 Å for

any interatomic distance between two protein molecules. This process was repeated until all the hard spheres in the simulation box were successful replaced by protein molecules. The number, N, of crowder molecules in the simulation boxes ranged from 6 to 48, in increments of 6. At the highest number, the crowder concentrations were 217 mg/mL for LYS and 196 mg/mL for BSA.

FMAP entails fictitiously placing a test protein into the crowder box and calculating the interaction energy between the test protein and the crowder proteins. The interaction energy function is given by equations (1)–(3), and the parameters are those described in our previous study. Specifically, the Lennard-Jones parameters were taken from Autodock (Morris et al., 2009) and the partial charges were taken from Amber (Cornell et al., 1995). To achieve a better balance between U_{n-a} and U_{elec} (as judged by, e.g., salt and temperature dependences of second virial coefficients of proteins in unpublished work), we scaled down the former by a factor of 0.2 (for comparison, Autodock applied a scaling factor around 0.16), and scaled up the latter by a factor of 2.0. The temperature was 298 K (where the solvent dielectric constant was 78.4) and the ionic strength was 0.15 M.

At each crowder concentration, 10 independent configurations of crowders were generated; for each crowder configuration, 500 random orientations for each test protein were chosen. So altogether 5,000 FMAP calculations were carried out for each test protein at each crowder concentration, and the results were averaged to yield the transfer free energy. To test the additivity between the excluded-volume component and the

soft-attraction component, we also carried out corresponding averaging to obtain these components (see below).

Two methods were used to do error analysis. The first was bootstrap. Here the 5,000 individual values of the transfer free energy (or a component thereof) were pooled to create the original sample. Bootstrap samples (with the same size, 5,000, as in the original sample) were then generated by randomly drawing from the original sample, and the standard deviation of the data in each bootstrap sample was calculated. The error of the FMAP calculation was finally estimated as the mean of the standard deviations of 10,000 bootstrap samples. The second was the block decorrelation technique of Flyvbjerg and Petersen (1989) (code downloaded from https://github.com/manoharan-lab/flyvbjerg-std-err/). Here we treated the 5,000 data points as a time series. These data points were "blocked" in successive generations. Specifically, the data points in the first generation were the original ones; in the second generation, the first two data points, the next two data points, and so on were each "blocked," i.e., merged and replaced by their averages. This blocking process continued until the total number of blocked data points went below a cutoff of 15. At each generation, the variance of the blocked data points was calculated. The variance reached a plateau before the cutoff, and the square root of the plateau value was taken as the error estimate.

RESULTS

Additivity Between Excluded-Volume and Soft-Attraction Components

As shown by equation (4), the transfer free energy $\Delta\mu$ is given by the average of the Boltzmann factor of the protein-crowder interaction energy U_{int}; the average needs to be taken over the position $\mathbf{R}$ of a fictitious placement of the test protein into the crowder box, the orientation Ω of the test protein, and the configuration $\mathbf{c}$ of the crowders. For a given Ω and a given $\mathbf{c}$, FMAP calculates the average over $\mathbf{R}$ from values of U_{int} at grid points within the crowder box. The grid points can be separated into ones with protein-crowder clash and ones that are clash-free. Note that $\exp\left(-U_{\text{st}}/k_B T\right)$ has value 0 at the clashed grid points and value 1 at the clash-free ones; the two soft interactions only operate at the clash-free grid points. Based on these considerations, we can write the average of $\exp\left(-U_{\text{int}}/k_B T\right)$ over $\mathbf{R}$ as

$$< \exp\left(-U_{\text{int}}/k_B T\right) >_{\mathbf{R}} = < \exp\left(-U_{\text{st}}/k_B T\right) >_{\mathbf{R}} \quad (5)$$
$$\times < \exp\left[-(U_{\text{n}-\text{a}} + U_{\text{elec}})/k_B T\right] >_1$$

where $< \cdots >_{\mathbf{R}}$ and $< \cdots >_1$ signify averaging over all the grid points and clash-free ones, respectively. Note that $< \exp\left(-U_{\text{st}}/k_B T\right) >_{\mathbf{R}}$ is simply the clash-free fraction of grid points. Corresponding to the factorization in equation (5), we can write the transfer free energy, calculated without averaging over Ω and $\mathbf{c}$, as the sum of an excluded-volume component and a soft-attraction component:

$$\Delta\mu = \Delta\mu_{\text{e}-\text{v}} + \Delta\mu_{\text{s}-\text{a}}, \text{for a single } \Omega \text{ and a single } \mathbf{c} \quad (6)$$

The excluded-volume component is given by the clash-free fraction,

$$\exp\left(-\Delta\mu_{\text{e}-\text{v}}/k_B T\right) = < \exp(-U_{\text{st}}/k_B T) >_{\mathbf{R}} \quad (7)$$

whereas the soft-attraction component is given by the combined soft interactions at the clash-free grid points:

$$\exp\left(-\Delta\mu_{\text{s}-\text{a}}/k_B T\right) = < \exp\left[-(U_{\text{n}-\text{a}} + U_{\text{elec}})/k_B T\right] >_1 \quad (8)$$

We further averaged $< \exp\left(-U_{\text{int}}/k_B T\right) >_{\mathbf{R}}$ over combinations of 500 test protein orientations and 10 crowder configurations to obtain the transfer free energy $\Delta\mu$. Specifically, the algebraic average of 5,000 individual values of $< \exp\left(-U_{\text{int}}/k_B T\right) >_{\mathbf{R}}$ was calculated and then converted to $\Delta\mu$. Similarly, we averaged $< \exp(-U_{\text{st}}/k_B T) >_{\mathbf{R}}$ and $< \exp\left[-(U_{\text{n}-\text{a}} + U_{\text{elec}})/k_B T\right] >_1$ over the 5,000 $\Omega/\mathbf{c}$ combinations to obtain $\Delta\mu_{\text{e}-\text{v}}$ and $\Delta\mu_{\text{s}-\text{a}}$, respectively. The sum of $\Delta\mu_{\text{e}-\text{v}}$ and $\Delta\mu_{\text{s}-\text{a}}$ provides a very accurate estimate of $\Delta\mu$ (**Figure 2**), demonstrating the additivity of these two components. The errors reported by two methods are very similar (**Figure S1**), and hence in **Figure 2** and hereafter we only show errors determined by the bootstrap method.

If we have an infinitely large crowder box, then its different regions give a good representation for the configurations of a finite crowder box. Likewise, when the test protein with a single orientation is fictitiously placed into different regions of the infinite crowder box, it is as if many different orientations of the test protein are probed by a finite crowder box. Hence, for an infinitely large crowder box, a separate average over crowder configurations and test protein orientations is unnecessary; then the separation of $\Delta\mu$ into $\Delta\mu_{\text{e}-\text{v}}$ and $\Delta\mu_{\text{s}-\text{a}}$ is exact. That our results accurately conform to additivity provides an indication that our crowder box is sufficiently large. In particular, the clash-free fraction, $< \exp(-U_{\text{st}}/k_B T) >_{\mathbf{R}}$, is highly constant among the 5,000 $\Omega/\mathbf{c}$ combinations, as indicated by very small $\Delta\mu_{\text{e}-\text{v}}$ errors (<0.01 kcal/mol, except for polε in the most concentrated LYS solution, where the error is 0.02 kcal/mol; see also below)
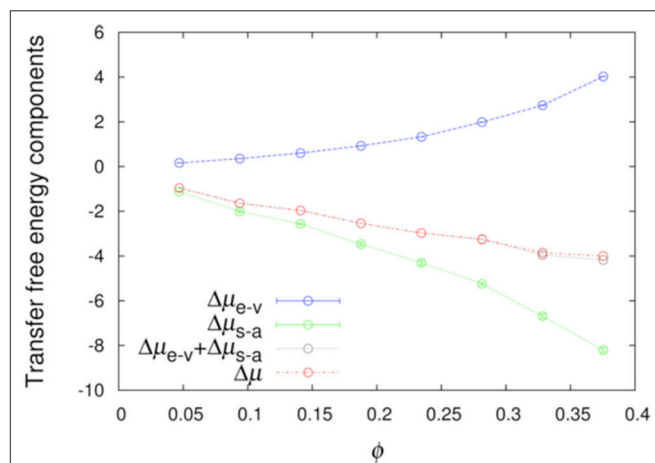

FIGURE 2 | Additivity of the excluded-volume and soft-attraction components of the transfer free energy.

(**Figure 2**). The decomposition into $\Delta\mu_{e-v}$ and $\Delta\mu_{s-a}$ provides physical insight into the transfer free energy. $\Delta\mu_{e-v}$ is necessarily positive, whereas $\Delta\mu_{s-a}$ can be expected to be negative. As **Figure 2** shows, these two quantities largely cancel each other, leading to a relatively modest magnitude for $\Delta\mu$.

We also asked whether $\Delta\mu_{s-a}$ could be further decomposed into the separate contributions of the two types of soft interactions. To that end, we carried out the averages of $\exp\left(-U_{n-a}/k_B T\right)$ and $\exp\left(-U_{elec}/k_B T\right)$, i.e., by including only one of the two types of soft interactions. For convenience we refer to the corresponding chemical potentials as $\Delta\mu_{n-a}$ and $\Delta\mu_{elec}$, respectively. As shown in **Figure S2**, the magnitude of μ_{s-a} is much larger than the sum of $\Delta\mu_{n-a}$ and $\Delta\mu_{elec}$, indicating strong correlations between the two types of soft interactions. Indeed, one expects that the strongest electrostatic attractions occur when the test protein is apposed to crowder proteins with high charge and shape complementarity, but high shape complementarity also leads to strong non-polar attraction.

For the largest test protein, polε, in the most concentrated LYS solution (217 mg/mL), the clash-free fraction in 2,086 of the 5,000 $\Omega/\mathbf{c}$ combinations was 0, i.e., not a single grid point was clash-free. In this concentrated crowder solution, the probability that voids large enough to accommodate polε is very small, which explains the high percentage of fully clashed $\Omega/\mathbf{c}$ combinations as well as the relatively higher error of $\Delta\mu_{e-v}$ (calculated on the 2,914 $\Omega/\mathbf{c}$ combinations with clash-free grid points). The $\Delta\mu_{e-v}$ value thus calculated was corrected by adding $-k_B T \ln\left(2,914/5,000\right)$ to account for the fully clashed $\Omega/\mathbf{c}$ combinations. The same correction also applies to $\Delta\Delta\mu$ in this case. Values of $\Delta\mu$, μ_{e-v}, and $\Delta\mu_{s-a}$ for the eight test proteins are presented in **Table S1** for LYS crowding and **Table S2** for BSA crowding.

Theoretical Modeling of Excluded-Volume Component

For calculating the excluded-volume transfer free energy, scaled-particle and other theories have been developed for test particles and crowder particles that have spherical and other simple shapes. Our generalized fundamental measure theory (GFMT) has enabled the test proteins to be represented at the all-atom level, though crowders still have to be modeled as spheres (Qin and Zhou, 2010). GFMT predicts the excluded-volume component as

$$\Delta\mu_{e-v} = \Pi_c v_p + \gamma_c s_p + \kappa_c l_p - k_B T \ln\left(1 - \phi\right) \quad (9)$$

where v_p, s_p, and l_p are the volume, surface area, and integrated mean curvature (with dimension of length) of the test protein; Π_c is the osmotic pressure of the crowder solution, and γ_c and κ_c are the corresponding quantities for surface tension and bending rigidity; and ϕ is the total volume fraction of the crowders. The latter is given by $\phi = V_c \rho_c$, where V_c and ρ_c are the volume and number density of the crowders. Two other quantities, $\rho_R = R_c \rho_c$ and $\rho_S = S_c \rho_c$, with R_c and S_c denoting the radius and surface

area of the crowders, are needed to define Π_c, γ_c, and κ_c. The results are

$$\frac{\Pi_c}{k_B T} = \frac{\rho}{1 - \phi} + \frac{\rho_R \rho_S}{(1 - \phi)^2} + \frac{\rho_S^3}{12\pi(1 - \phi)^3} \quad (10)$$

$$\frac{\gamma_c}{T} = \frac{\rho_R}{1 - \phi} + \frac{\rho_S^2}{8\pi(1 - \phi)^2} \quad (11)$$

$$\frac{\kappa_c}{k_B T} = \frac{\rho_S}{1 - \phi} \quad (12)$$

The osmotic pressure can be viewed as the energy to create a cavity with a unit volume in the crowder solution; the surface tension is the energy to create a unit-area interface between the crowder solution and a test protein; and the bending rigidity measures the energy arising from the curvature of the interface. Fitting our $\Delta\mu_{e-v}$ data to GFMT meant that we modeled the crowder proteins as spheres; in so doing we needed to specify the radius, R_c, for each crowder protein. Note that R_c is the only free parameter; once R_c is chosen, the volume, surface area, and linear size of the test protein (i.e., v_p, s_p, and l_p) are calculated by rolling a spherical probe of radius R_c around the three-dimensional structure of the test protein.

We were able to achieve a good global fit for all the eight test proteins in either LYS or BSA after searching for an R_c value that minimized deviations between the $\Delta\mu_{e-v}$ data and the GFMT predictions (**Figure 3**). The resulting R_c values are 21.4 and 35.4 Å, respectively, for LYS and BSA. These values are close to the hydrodynamic radii, 19.6 and 36.5 Å, calculated by HYDROPRO (Ortega et al., 2011). Using the preceding R_c values, the volume fractions of the two crowders at their highest concentrations are 37.6% and 33.1%. The resulting values for v_p, s_p, and l_p of the eight test proteins are presented in **Table S3**.

Quadratic Fitting of Soft-Attraction Component

The soft-attraction component, $\Delta\mu_{s-a}$, calculated by FMAP has a nearly linear dependence on the crowder concentration (**Figure 4**). We fitted the results to a quadratic function:

$$\Delta\mu_{s-a} = a\phi + b\phi^2 \quad (13)$$

The quadratic term makes a minor contribution in most of the 16 sets of results (eight test proteins pairs with two crowder proteins). In particular, at $\phi = 30\%$, the quadratic term is $<20\%$ of the linear term in 9 out of the 16 cases.

Importantly, when the data at the first seven crowder concentrations were used for the fitting, the fitting function, by extrapolation, predicts well the $\Delta\mu_{s-a}$ result at the eighth crowder concentration (comparing the red circle and the eighth blue circle in each case). This suggests that a function like equation (15) can be used to predict $\Delta\mu_{s-a}$ at high crowder concentrations, where FMAP calculations become difficult because voids that can accommodate the test proteins are rare.
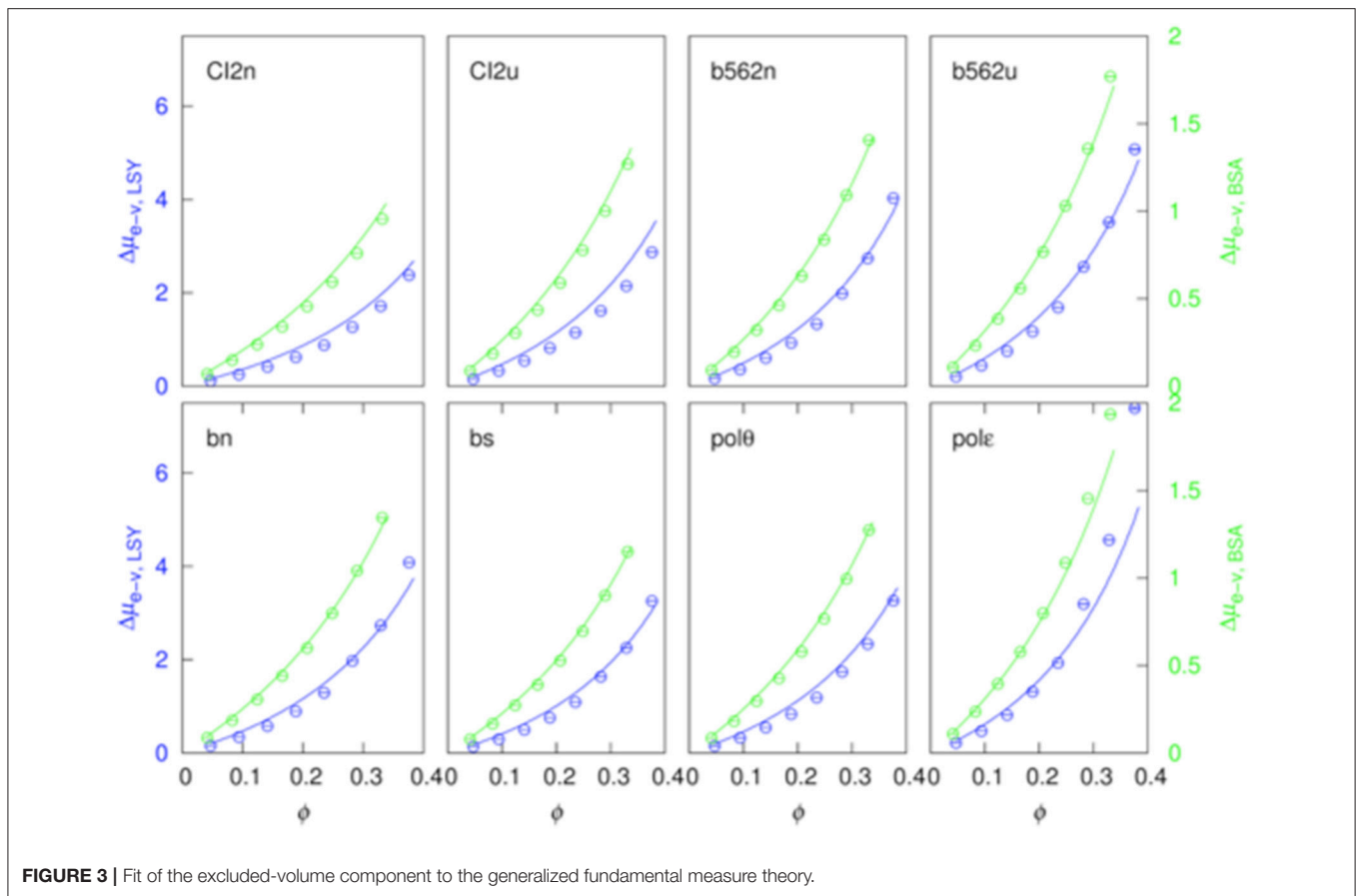
FIGURE 3 | Fit of the excluded-volume component to the generalized fundamental measure theory.

DISCUSSION

By using our FMAP, we have calculated the transfer energies of eight test proteins over a wide range of concentrations of two crowder proteins. We have shown that the transfer free energy can be accurately decomposed into an excluded-volume component, arising from the hard-core repulsion, and a soft-attraction component, arising from non-polar and electrostatic interactions. Our calculation results thus rigorously validate similar decompositions proposed previously (Petsev et al., 2003; Jiao et al., 2010; Minton, 2013). We have found that the excluded-volume component is predicted well by the generalized fundamental measure theory, which was developed for atomistic test proteins in the presence of spherical crowders that exert only steric repulsion. On the other hand, we have found that the soft-attraction component has a nearly linear dependence on crowder concentration. The latter result is interesting and has important implications.

Why does $\Delta\mu_{s-a}$ have a nearly linear dependence on crowder concentration? To gain insight, we turn to the perturbed virial expansion for pure molecular fluids (Nezbeda and Smith, 2004). The expansion was originally applied to the pressure (P),

$$\frac{P}{k_B T \rho} = \frac{P_{\text{ref}}}{k_B T \rho} + \sum_{l \geq 2} \Delta B_l \rho^{l-1} \tag{14}$$
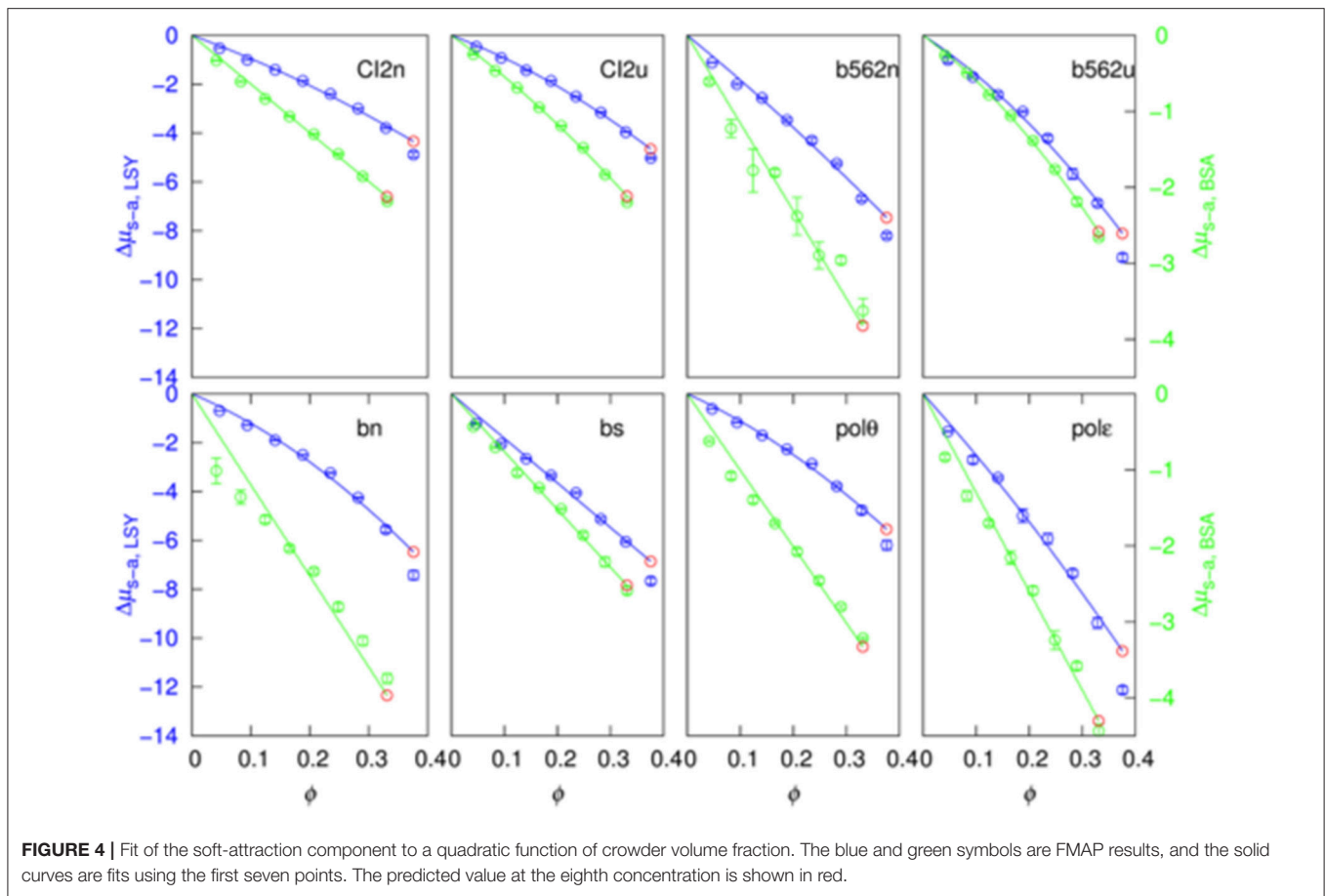
where ρ is the number density, P_{ref} is the pressure of a reference system, and ΔB_l are the "residual" virial coefficients, i.e., the differences in virial coefficients between the real and reference system. We can easily turn equation (14) into an expression for the excess chemical potential, using the relation (Qin and Zhou, 2016)

$$\mu^{\text{ex}} = \int_0^\rho \frac{1}{\rho} \frac{\partial P}{\partial \rho} d\rho \tag{15}$$

The result is

$$\mu^{\text{ex}} = \mu^{\text{ex}}_{\text{ref}} + k_B T \sum_{l \geq 2} [l/(l-1)] B_l \rho^{l-1} \tag{16}$$

A protein-crowder system where the test protein and crowder protein are the same is equivalent to a pure molecular fluid. In that case, the transfer free energy $\Delta\mu$ and the excess chemical potential μ^{ex} are equivalent. Furthermore, we may choose the reference system such that $\mu^{\text{ex}}_{\text{ref}}$ is equivalent to $\Delta\mu_{e-v}$, then the second term, an infinite sum and to be denoted as $\Delta\mu^{\text{ex}}$, on the right-hand side of equation (16) is equivalent to $\Delta\mu_{s-a}$. We can now recognize the quadratic function in equation (13) as a truncation of the infinite sum to the second order. The nearly linear dependence $\Delta\mu_{s-a}$ on crowder concentration just means

FIGURE 4 | Fit of the soft-attraction component to a quadratic function of crowder volume fraction. The blue and green symbols are FMAP results, and the solid curves are fits using the first seven points. The predicted value at the eighth concentration is shown in red.

that the contributions of the second and higher orders are much less than that of the first order.

In **Figure S3**, we display the contributions of the first-, second-, and third-order contributions to $\Delta\mu^{\text{ex}}$ for Lennard-Jones fluids, with the reference system chosen as hard-sphere fluids (diameter $= \sigma$). Over a wide range of ε, the depth of the interaction potential, the dominant contribution comes from the first order. Virial coefficients B_l are integrals of Mayer functions over the positions of l molecules. A Mayer function is the Boltzmann factor of the intermolecular interaction potential subtracted by 1; hence we expect that dominant contributions to residual virial coefficients ΔB_l come from clusters of l molecules in which all pairs are in the most attractive range of intermolecular distance. When the range of attraction is narrow, such molecular clusters become rare for $l = 3$ and higher. That would lead to small $\Delta B_{l \geq 3}$ values and explain why $\Delta\mu^{\text{ex}}$ is dominated by the first-order term, which is proportional to ΔB_2. The foregoing argument for pure molecular fluids largely applies to the protein-crowder systems studied in the present work, thus providing a rationalization for the nearly linear dependence of $\Delta\mu_{\text{s}-\text{a}}$ on crowder concentration. An interesting future study would be to directly validate this argument by calculating virial coefficients of different orders for protein-crowder systems. It is straightforward to apply FMAP for B_2 calculations, but efficient $B_{l \geq 3}$ calculations for protein-crowder systems will require careful algorithmic

design. Minton modeled soft attraction as weak unsaturable binding (Minton, 2013), which leads to an approximately linear dependence on ϕ. Hoppe and Minton (Hoppe and Minton, 2016) used a perturbed virial expansion, similar to equation (16) and including ΔB_2 and ΔB_3, for square-well crowders and found $\Delta\mu_{\text{s}-\text{a}}$ to be linearly dependent on ϕ. The present work provides confirmation of these previous results and generalize them to atomistic models.

A nearly linear dependence of $\Delta\mu_{\text{s}-\text{a}}$ allows us to extrapolate results obtained at lower crowder concentrations to higher ones, as we demonstrated here (**Figure 4**). As noted above, calculation of transfer free energies at high crowder concentrations becomes challenging for FMAP and likewise for other methods. For FMAP, the high percentage of fully clashed $\Omega/\mathbf{c}$ combinations for the largest test protein studied here in the most concentrated LYS solution gives an indication of this challenge. A similar situation was observed in our previous study of disordered proteins in the presence of polyethylene glycol (Nguemaha et al., 2018). Extrapolation to higher crowder concentrations may provide a means to finesse this challenge.

While our interaction potential is atomistic, it is based on implicit solvent modeling. As a result, the treatment of hydration effects may prove inadequate, thereby limiting the accuracy of model predictions. In future studies we will investigate the performance of our atomistic model in quantitatively predicting experimental observables, including second virial coefficients.

As a final note, for determining the liquid-liquid phase equilibria of protein (Qin and Zhou, 2016) and colloid (Lomakin et al., 1996) solutions, it has been found that at least a third-order fitting is needed for the soft-attraction component of the excess chemical potential. In that case, a precise dependence of the chemical potential on protein concentration is key to determining the phase boundary, and hence one must calculate the chemical potential over as wide a range of protein concentration and cover the concentration range with as many points as can be done. For our protein-crowder systems, we find that, in a cubic fit, the second and third order terms have opposite signs and hence largely cancel each other, still leaving the first-order-term dominant.

AUTHOR CONTRIBUTIONS

VN conducted research and analyzed data. SQ prepared methods and engaged in discussion. H-XZ supervised research and wrote the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2019.00039/full#supplementary-material

Three figures comparing two methods of error estimates (**Figure S1**), showing non-additivity of the contributions to $\Delta\mu_{s-a}$ by the two types of soft interactions (**Figure S2**), and illustrating relatively small contributions to $\Delta\mu_{s-a}$ by higher order terms in a perturbed virial expansion of $\Delta\mu^{ex}$ (**Figure S3**); and three supporting tables, presenting data on transfer free energies (**Tables S1,S2**) and geometric properties (**Table S3**) of test proteins.

REFERENCES

Banks, A., Qin, S. K., Weiss, L. C., Stanley, B., and Zhou, H. X. (2018). Intrinsically disordered protein exhibits both compaction and expansion under macromolecular crowding. *Biophys. J.* 114:1067–1079. doi: 10.1016/j.bpj.2018.01.011

Batra, J., Xu, K., Qin, S., and Zhou, H. X. (2009b). Effect of macromolecular crowding on protein binding stability: modest stabilization and significant biological consequences. *Biophys. J.* 97:906–911. doi: 10.1016/j.bpj.2009.05.032

Batra, J., Xu, K., and Zhou, H. X. (2009a). Nonaddtive effects of mixed crowding on protein stability. *Proteins* 77:133–138. doi: 10.1002/prot.22425

Cheung, M. S., Klimov, D., and Thirumalai, D. (2005). Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4753–4758. doi: 10.1073/pnas.0409630102

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., et al. (1995). A Second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117, 5179–5197. doi: 10.1021/ja00124a002

Flyvbjerg, H., and Petersen, H. G. (1989). Error-estimates on averages of correlated data. *J. Chem. Phys.* 91, 461–466. doi: 10.1063/1.457480

Gnutt, D., and Ebbinghaus, S. (2016). The macromolecular crowding effect–from *in vitro* into the cell. *Biol. Chem.* 397:37–44. doi: 10.1515/hsz-2015-0161

Goldenberg, D. P., and Argyle, B. (2014). Self crowding of globular proteins studied by small-angle x-ray scattering. *Biophys. J.* 106, 895–904. doi: 10.1016/j.bpj.2013.12.004

Hoppe, T., and Minton, A. P. (2016). Incorporation of hard and soft protein-protein interactions into models for crowding effects in binary and ternary protein mixtures. Comparison of approximate analytical solutions with numerical simulation. *J. Phys. Chem. B*, 120, 11866–11872. doi: 10.1021/acs.jpcb.6b07736

Jiao, M., Li, H. T., Chen, J., Minton, A. P., and Liang, Y. (2010). Attractive protein-polymer interactions markedly alter the effect of macromolecular crowding on protein association equilibria. *Biophys. J.* 99, 914–923. doi: 10.1016/j.bpj.2010.05.013

Kim, Y. C., and Mittal, J. (2013). Crowding induced entropy-enthalpy compensation in protein association equilibria. *Phys. Rev. Lett.* 110:208102. doi: 10.1103/PhysRevLett.110.208102

Lomakin, A., Asherie, N., and Benedek, G. B. (1996). Monte Carlo study of phase separation in aqueous protein solutions. *J. Chem. Phys.* 104:1646–1656. doi: 10.1063/1.470751

McGuffee, S. R., and Elcock, A. H. (2010). Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput Biol.* 6:e1000694. doi: 10.1371/journal.pcbi.1000694

Miklos, A. C., Sarkar, M., Wang, Y., and Pielak, G. J. (2011). Protein crowding tunes protein stability. *J Am Chem Soc.* 133, 7116–7120. doi: 10.1021/ja200067p

Miklos, A. C., Sumpter, M., and Zhou, H. X. (2013). Competitive interactions of ligands and macromolecular crowders with maltose binding protein. *PLoS ONE* 8:e74969. doi: 10.1371/journal.pone.0074969

Minh, D. D., Chang, C. E., Trylska, J., Tozzini, V., and McCammon, J. A. (2006). The influence of macromolecular crowding on HIV-1 protease internal dynamics. *J. Am. Chem. Soc.* 128, 6006–6007. doi: 10.1021/ja060483s

Minton, A. P. (1981). Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers*. 20, 2093–2120. doi: 10.1002/bip.1981.360201006

Minton, A. P. (1983). The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol. Cell Biochem.* 55, 119–140. doi: 10.1007/BF00673707

Minton, A. P. (2013). Quantitative assessment of the relative contributions of steric repulsion and chemical interactions to macromolecular crowding. *Biopolymers*. 99, 239–244. doi: 10.1002/bip.22163

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256

Nezbeda, I., and Smith, W. R. (2004). On the calculation of the critical temperature from the second virial coefficient. *Fluid Phase Equil.* 216, 183–186. doi: 10.1016/j.fluid.2003.11.006

Nguemaha, V., Qin, S., and Zhou, H. X. (2018). Atomistic modeling of intrinsically disordered proteins under polyethylene glycol crowding: quantitative comparison with experimental data and implication of protein-crowder attraction. *J. Phys. Chem. B*. 122, 11262–11270. doi: 10.1021/acs.jpcb.8b07066

Ortega, A., Amorós, D., and García de la Torre, J. (2011). Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J*. 101, 892–898. doi: 10.1016/j.bpj.2011.06.046

Petsev, D. N., Wu, X. X., Galkin, O., and Vekilov, P. G. (2003). Thermodynamic functions of concentrated protein solutions from phase equilibria. *J. Phys. Chem. B*. 107, 3921–3926. doi: 10.1021/jp0278317

Phillip, Y., Harel, M., Khait, R., Qin, S., Zhou, H. X., and Schreiber, G. (2012). Contrasting factors on the kinetic path to protein complex formation diminish the effects of crowding agents. *Biophys. J.* 103, 1011–1019. doi: 10.1016/j.bpj.2012.08.009

Qin, S., Pang, X., and Zhou, H. X. (2011). Automated prediction of protein association rate constants. *Structure* 19, 1744–1751. doi: 10.1016/j.str.2011.10.015

Qin, S., and Zhou, H. X. (2009). Atomistic modeling of macromolecular crowding predicts modest increases in protein folding and binding stability. *Biophys. J.* 97, 12–19. doi: 10.1016/j.bpj.2009.03.066

Qin, S., and Zhou, H. X. (2010). Generalized fundamental measure theory for atomistic modeling of macromolecular crowding. *Phys. Rev. E.* 81:031919. doi: 10.1103/PhysRevE.81.031919

Qin, S., and Zhou, H. X. (2013). An FFT-based method for modeling protein folding and binding under crowding: benchmarking on ellipsoidal and all-atom crowders. *J. Chem. Theory Comput.* 9:10. doi: 10.1021/ct4005195

Qin, S., and Zhou, H. X. (2014). Further development of the FFT-based method for atomistic modeling of protein folding and binding under crowding: optimization of accuracy and speed. *J. Chem. Theory Comput.* 10, 2824–2835. doi: 10.1021/ct5001878

Qin, S., and Zhou, H. X. (2016). Fast method for computing chemical potentials and liquid-liquid phase equilibria of macromolecular solutions. *J. Phys. Chem. B.* 120, 8164–8174. doi: 10.1021/acs.jpcb.6b01607

Rivas, G., Fernandez, J. A., and Minton, A. P. (1999). Direct observation of the self-association of dilute proteins in the presence of inert macromolecules at high concentration via tracer sedimentation equilibrium: theory, experiment, and biological significance. *Biochemistry* 38, 9379–9388. doi: 10.1021/bi990355z

Sarkar, M., Smith, A. E., and Pielak, G. J. (2013). Impact of reconstituted cytosol on protein stability. *Proc. Natl. Acad. Sci. U.S.A.* 110, 19342–19347. doi: 10.1073/pnas.1312678110

Skoge, M., Donev, A., Stillinger, F. H., and Torquato, S. (2006). Packing hyperspheres in high-dimensional Euclidean spaces. *Phys. Rev. E.* 74:041127. doi: 10.1103/PhysRevE.74.041127

Soranno, A., Koenig, I., Borgia, M. B., Hofmann, H., Zosel, F., Nettels, D., et al. (2014). Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4874–4879. doi: 10.1073/pnas.1322611111

Wang, Y., Sarkar, M., Smith, A. E., Krois, A. S., and Pielak, G. J. (2012). Macromolecular crowding and protein stability. *J. Am. Chem. Soc.* 134, 16614–16618. doi: 10.1021/ja305300m

Wu, D., and Minton, A. P. (2013). Quantitative characterization of the interaction between sucrose and native proteins via static light scattering. *J. Phys. Chem. B.* 117, 111–117. doi: 10.1021/jp308880v

Zhou, H. X. (2013). Influence of crowded cellular environments on protein folding, binding, and oligomerization: biological consequences and potentials of atomistic modeling. *FEBS Lett.* 587, 1053–1061. doi: 10.1016/j.febslet.2013.01.064

Zhou, H. X., Rivas, G., and Minton, A. P. (2008). Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* 37, 375–397. doi: 10.1146/annurev.biophys.37.032807.125817