2014

Effect of Macromolecular Crowding on the Electrostatic Interaction of Barnase-Barstar: Initial Steps Using an Explicit Solvent Model

Helena Qi
hqi@wellesley.edu

Effect of Macromolecular Crowding on the Electrostatic Interaction of Barnase-Barstar: Initial Steps Using an Explicit Solvent Model

Helena Qi

Advisor: Mala Radhakrishnan

April 2014

Submitted in Partial Fulfillment of the Prerequisites for Honors in Chemistry

Abstract

Proteins that bind intracellularly are surrounded by other macromolecules. Macromolecular crowding has been shown to impact protein folding and binding, but its effects on the electrostatics of protein binding have not been thoroughly studied. Two ways crowding can affect binding are via loss of water mobility and water depletion. Crowding causes loss of water mobility because more water molecules will be organized into solvation shells around the crowding agents, instead of being in bulk form; waters in solvation shells are less mobile. Water is depleted because the crowders occupy volume that would have held water. We are interested in the effect of loss of water mobility, but it is difficult to separate the effects of loss of water mobility and water depletion. We had previously used implicit solvent to study the effect of water depletion due to crowding. An explicit solvent model will show both loss of water mobility and water depletion effects. In combination with the results from the implicit solvent study, we can then assess the effect of loss of water mobility. In this study, we used free energy perturbation and component analysis in explicit water to begin to examine the binding of barnase and barstar. Specifically, we evaluated the contribution of the charge of a particular residue, barstar's aspartic acid 35, to the binding free energy of the barnase-barstar complex. In the future, we will introduce crowders into the system, so we may see how the contribution of the charge of a residue to the binding free energy changes in the presence of crowders.

Acknowledgements

First and foremost, I am grateful for my advisor, Mala Radhakrishnan, whose guidance and support through the last three and a half years has been invaluable for my development as a scientist and researcher. To live up to the expectations she sets has pushed me to deeper understanding and better presentations.

Second, I would like to thank my thesis committee members, Professors Donald Elmore, Christopher Arumainayagam, and Glenn Stark, who have helped me on the path to this thesis. In particular, Professor Elmore's guidance in using GROMACS has been indispensable.

Third, my fellow Radhakrishnan lab mates have been a constant source of encouragement and were generally the people I turned to first when I needed to discuss my research. Connie Chen '15 and Priyanka Nakka '12 helped develop the crowding model used in this thesis. Thanks also to Emma Nechamkin '12, who showed me the ropes when I first started out.

Fourth, my fellow thesising seniors, friends, and family who commiserated with me when code was buggy and celebrated when it ran. Thanks especially to Hannah Herde '14, with whom I have taken every physics class at Wellesley and who helped edit part of this thesis.

Fifth, this work would not have been possible without the Roberta Day Staley and Karl A. Staley Fellowship for Cancer Related Research.

Contents

1 Introduction

1.1 Crowding

Many of the current models of protein binding, vital to rational drug design and to understanding the fundamentals of protein interaction, represent proteins and other species as surrounded by a sea of water containing salt. However, proteins in real cells are in a crowded environment. Between 5% and 40% of cells are composed of macromolecules,[1] which could potentially affect the binding and recognition of protein partners. For example, a theoretical study of relevance to anticancer strategies[2] showed that for ligands binding to telomeric DNA, the binding affinity of the ligands was significantly affected by molecular crowding. Crowding also changes the rates[3,4] and mechanisms[5,6] of biological processes. In order to understand protein binding within the cellular environment, it is necessary to understand how crowding affects binding.

There are two ways crowding can affect proteins – through hard excluded volume interactions and soft chemical ones. Many previous studies have focused on the entropic excluded volume effect, which is caused by crowders occupying volume that was once accessible to the protein.[7] When the space accessible to a protein decreases, the entropy of noncompact states decreases more than that of compact states, and the compact states (e.g. folded, bound, or aggregated) are thus relatively stabilized. For example, the crowding agents Ficoll and dextran, which have little soft chemical interaction with proteins, cause proteins to favor a compact state.[8,9]

Soft chemical interactions, unlike excluded volume effects, can be either stabilizing or destabilizing and mostly affect enthalpy. For example, urea and trimethylamine N-oxide (TMAO) both interact with the protein backbone, but urea destabilizes[10] while TMAO stabilizes it.[11] *In vitro*, it is difficult to separate the effects of soft and hard interactions. In a study using poly(vinylpyrrolidone), which interacts only weakly with proteins[12] and was deliberately chosen to minimize soft interactions, the soft interactions were still significant

enough to reduce the stabilization caused by volume exclusion.[13]

One particularly important physical determinant of molecular binding is electrostatics. However, the effect of crowding on the electrostatic component of binding has not been thoroughly studied. Previous studies[14–16] have mainly focused on how protein conformation is affected, leaving the effect on binding free energy largely unexplored.

1.2 Electrostatics

Many important aspects of protein binding, such as binding affinity, specificity, and promiscuity have been shown to be dependent on electrostatics.[17] In particular, the interaction specificity of a molecule with its correct binding partner is important not only for understanding living cells, but also for targeting specific molecules with drugs. Electrostatics have been used to design mutations on proteins[18] and modifications of drugs[19] with different affinities and specificities. These studies all model the proteins and drugs as surrounded by a sea of water, without any crowding.

Water in biological systems can be thought of as occurring in two phases: bulk water and hydration shells. Water molecules will organize themselves into solvation shells around other molecules. In the form of hydration shells, water is less mobile than if it were in bulk form. These non-bulk effects can extend a few Angstroms to several nanometers from the biomolecular-solvent interface.[20] The amount of non-bulk water in a cell is controversial. Magnetic relaxation dispersion suggests 20% of cell water content is non-bulk, which corresponds only to the first hydration shell.[20] X-ray and neutron scattering[21] and diffusion of endogenous probes[22–24] suggests that up to 75% of cell water may be non-bulk.

Crowding can decrease the mobility of water.[25] When more crowding agents are present, more water molecules will be in the form of solvation shells instead of bulk water. Additionally, by replacing water molecules with less polarizable crowding agents, electrostatic interactions between the protein partners are descreened. Previous studies have incorporated crowding effects into a continuum model of solvation,[26] and a large-scale simulation of

a crowded system has partially accounted for electrostatic solvation effects.[27] Computationally, crowders in solvent appear to lower the effective dielectric constant of the solvent;[14–16] this agrees with experiments that measure a lower dielectric constant in the cytoplasm, as compared to that of pure water.[28,29]

In a previous study carried out in collaboration with Connie Chen '15 and Priyanka Nakka '12, we examined the effect of crowding on the electrostatic component of protein interaction in implicit solvent (i.e. the water is modeled as a constant dielectric).[30] The protein complex used, barnase-barstar, is known to have strong electrostatic interactions[31,32] and has been used in previous crowding studies.[26] In that study, we were concerned solely with the water depletion effect and not with altered water mobility. By modeling the water implicitly, we were able to solely examine the water depletion effect. To solve for the free energies, the Poisson equation (1) was used to relate the potential $\phi(\vec{r})$ to the charge distribution $\rho(\vec{r})$ and the dielectric $\epsilon(\vec{r})$. The crowders were modeled as uncharged spheres with the same dielectric constant as the protein; in effect, they were spherical cavities of lower dielectric within a high dielectric solvent. We found that the presence of crowders caused the protein partners, which are electrostatically optimized to bind, to bind more tightly, with larger average effects at higher crowder concentration and smaller crowder size.

$$\nabla \cdot (\epsilon(\vec{r})\nabla\phi(\vec{r})) = -\frac{\rho(\vec{r})}{\epsilon_0} \tag{1}$$

In this previous study, we also used a technique known as component analysis[32–36] to evaluate how the electrostatic contributions of specific amino acid residues were impacted by the crowders. Component analysis involves setting the charges on a residue or moiety to zero, and evaluating the change in binding free energy as shown in equation (2). A positive $\Delta\Delta G$ indicates the residue contributes favorably to binding. We can evaluate the effect of crowding on $\Delta\Delta G$ using (3). A positive $\Delta\Delta\Delta G$ indicates the residue contributes more favorably (i.e. is more important) or less unfavorably when there is crowding.

$$\Delta\Delta G = \Delta G_{zeroed} - \Delta G_{original} \tag{2}$$

$$\Delta\Delta\Delta G = \Delta\Delta G_{crowded} - \Delta\Delta G_{uncrowded} \tag{3}$$

Lee and Tidor[32] showed that the five barstar residues whose side chains contribute most to the electrostatic component of binding free energy are TYR29, ASP35, ASP39, THR42, and GLU76, shown in figure 1. Of these, we found that the charged side chains contributed even more favorably in the presence of crowders, though only on the order of tenths of a kcal/mol.[30]

The previous study focused on water depletion and therefore used an implicit solvent model. However, as previously discussed, there are also water mobility effects associated with crowding. In order to model water mobility, individual water molecules must be free to move, and so must be modeled explicitly. Molecular dynamics (MD), in combination with explicit water, has previously been used to study crowding.[14] However, it is very difficult to directly model a protein-protein binding process using MD, so in this study, we will combine component analysis with explicit solvent MD simulations to examine the effect of crowding on the contribution that particular residues have on the binding free energy.

1.3 Molecular Dynamics

Molecular dynamics (MD) follows the trajectory of atoms through time. MD treats the atoms as classical particles. Using empirical molecular mechanical force fields, described below, the forces acting upon the atoms can be determined. Once the forces are known, the motion of the atoms can be determined by numerically integrating Newton's equations of motion through time. If the system is ergodic and adequate sampling has been done, the time average over the simulation will be the same as the NPT ensemble average, allowing thermodynamic properties like free energy to be determined. It is often assumed, though not proven, that for a reasonably long simulation, the system is ergodic.

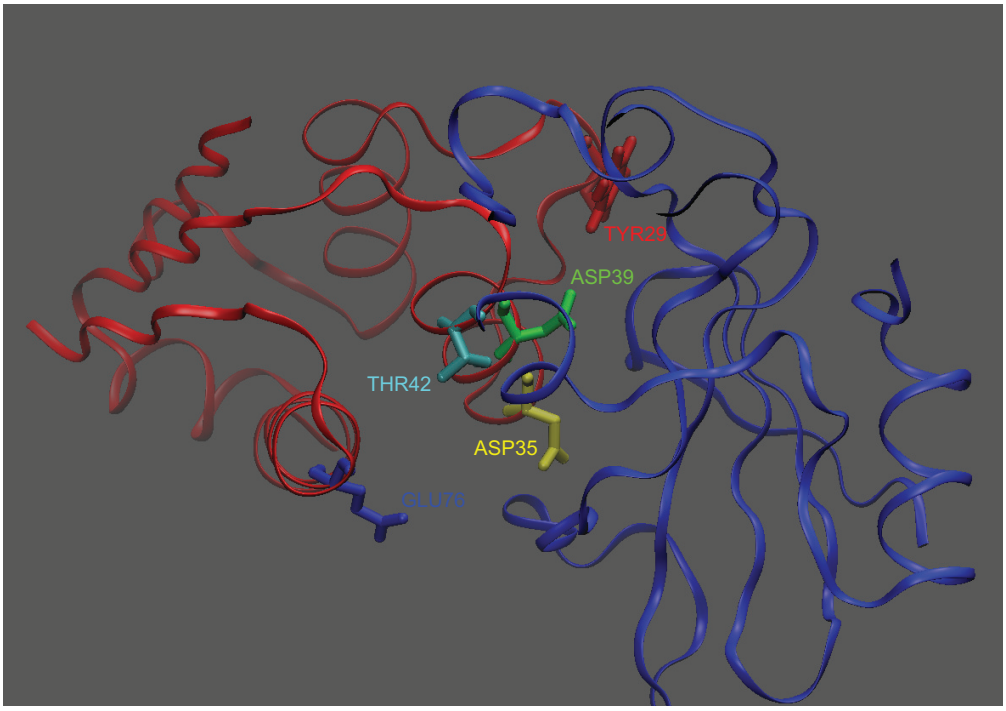


Figure 1: Five residues on barstar that contribute the most favorably towards the electrostatic component of barnase-barstar binding energy.

From Newtons laws, the position update (4) and velocity update (5) formulae can be derived. Here, $\vec{r}$ is position, $\vec{v}$ is velocity, $\vec{F}$ is force, t is time, and m is mass. Notice that if we are interested in the motion of a system, Δt must be short enough to capture these motions. Bond vibrations that involve hydrogen atoms occur on the order of femtoseconds; if we want to use a Δt larger than a fs, it is necessary to use an algorithm like LINCS,[37] which constrains bond lengths.

$$\Delta\vec{r} = \vec{v}_{avg}\Delta t \tag{4}$$

$$\Delta\vec{v} = \frac{\vec{F}_{avg}}{m}\Delta t \tag{5}$$

In order to integrate the position update (4) and velocity update (5) formulae, we use the second order leap-frog algorithm,[38] to produce equations (6) and (7). Leap-frog integration updates the position and velocity at staggered time points, with the positions defined on integer times and velocities defined on half-integer times.

$$\vec{r}(t) = \vec{r}(t - \Delta t) + \vec{v}(t - \frac{1}{2}\Delta t)\Delta t \tag{6}$$

$$\vec{v}(t + \frac{1}{2}\Delta t) = \vec{v}(t - \frac{1}{2}\Delta t) + \frac{\vec{F}(t)}{m}\Delta t \tag{7}$$

In order to run an MD simulation over a single time step, Δt, it is necessary to know the initial positions, velocities, and masses, in addition to the forces on each particle. In the case of the barnase-barstar complex, there is a known crystal structure to serve as a starting point.[39] The forces may be calculated once a potential energy (i.e. force field) is known, by taking the gradient of the energy.

Molecular mechanical force fields are derived from classical mechanical approximations to the quantum mechanical model of molecules, breaking down the potential energy into components, as seen in equation (8). Each of these components is parameterized from experiment or *ab initio* theoretical calculations.

$$E = E_{bond} + E_{angle} + E_{dihedral} + E_{electrostatic} + E_{vanderWaals} \tag{8}$$

The bond, angle, and dihedral terms are collectively known as the bonded terms and model interactions between atoms within a couple bonds of each other. E_{bond} is the energy associated with bond lengths deviating from their equilibrium values. E_{angle} is the energy associated with angles deviating from their equilibrium values. These two terms are often approximated by harmonic potentials. Notice that this model implies that no covalent bonds can be created or broken; to do so is beyond the scope of molecular mechanics. $E_{dihedral}$ is the energy associated with dihedral angles deviating from their equilibrium values, but it cannot be modeled simply as a harmonic oscillator; it is typically treated as a periodic function of the dihedral, with the periodicity dependent on the hybridization of the two middle atoms.

The van der Waals and electrostatic terms are collectively known as the non-bonded terms, and apply only to atoms that are not close together in bonded sequence. Recall that the van der Waals force is comprised of two effects, the attractive London dispersion force and

the repulsive Pauli exchange force; the van der Waals term will be interchangeably be called the Lennard-Jones (LJ) interaction in this thesis. The London dispersion force arises from attraction between induced dipole–induced dipole correlated electron motion. Specifically, the electron clouds surrounding the atoms have temporarily shifted such that the atoms are dipoles and these dipoles interact with each other. London dispersion falls off as $\frac{1}{r^6}$, where r is the distance between the two atoms, and so it is weak at long distances. The Pauli exchange repulsion is even shorter ranged and very strong. It arises from the fact that the wave functions of indistinguishable fermions must have exchange antisymmetry, and causes the steric effect. For computational ease, it is often modeled as proportional to $\frac{1}{r^{12}}$. Because the LJ potential decays so rapidly with r, it is typical to neglect its contribution at large r.

Of particular interest is the electrostatic term, which, as described previously, acts over long distances and has a large impact when modeling proteins. In reality, atoms are partially comprised of electron clouds, but for the sake of computational feasibility, they are not directly modeled. Instead, partial atomic point charges and atomic radii are assigned to the atoms. These parameters are determined through fitting to experiment and to *ab initio* calculations. Once the atoms have point charges, their electrostatic interactions can be modeled through Coulomb's law (9), noting that only atoms at least three bonds away from each other should be included. Here, q is the charge of the atom, r is the distance between two atoms, and k is Coulomb's constant.

$$E_{electrostatic} = \sum_{j \neq i} \sum_{i} \frac{k q_i q_j}{r_{ij}} \tag{9}$$

Notice that Coulomb's law is $O(N^2)$ in the number of atoms in the system. Unfortunately, because the electrostatic force drops off relatively slowly with distance, long-range electrostatic interaction is actually very important. Additionally, in an actual simulation, to avoid boundary issues, the boundary conditions are often periodic, so using Coulomb's law would naively take infinite time. One method to approximately calculate the electrostatic energy is the particle mesh Ewald (PME) method.[40] The PME method involves separating

the potential (in this case, electrostatic interaction potential) into short- and long-range components. The short-range sum converges quickly in real space, while the long-range component converges quickly in Fourier space. This way, PME can more rapidly compute interaction energies of periodic systems. The Fourier transform assumes periodicity, and so it neatly deals with the boundary condition used. Overall, PME gives results in $O(N\log N)$ time, where N is the number of atoms in each cell.

In cases where we desire the system to be coupled to a heat bath to maintain constant temperature, a Berendsen thermostat may be used.[41] The algorithm rescales velocities v according to equation (10) by coupling the temperature of the system T to a heat bath of temperature T_0 with coupling time constant τ. A similar Berendsen barostat is used to keep the pressure at a desired value; in that case, the positions are rescaled.

$$v' = v\sqrt{1 + \frac{\Delta t}{\tau}(\frac{T_0}{T} - 1)} \tag{10}$$

MD may be run with either explicit or implicit solvent. Because the goal here is to examine the effect of water mobility, we used explicit solvent, specifically the TIP3P water model.[42] TIP3P is a three-site model, with the three atoms in the water molecule able to interact with other atoms. The number of sites in the model indicates the number of potential sites of interaction. The molecule is held with a rigid geometry, so that the bond angle and lengths are constant, with each of the atoms having its own charge and Lennard-Jones parameters.

1.4 Free Energy Perturbation

The goal of the project is to calculate the $\Delta\Delta\Delta G$ for the five aforementioned important residues of barstar, shown in figure 1, in the presence of explicit solvent in order to assess the effect of crowding on the contributions these residues make to the binding free energy. Again, it is computationally infeasible to directly simulate the binding process. However,

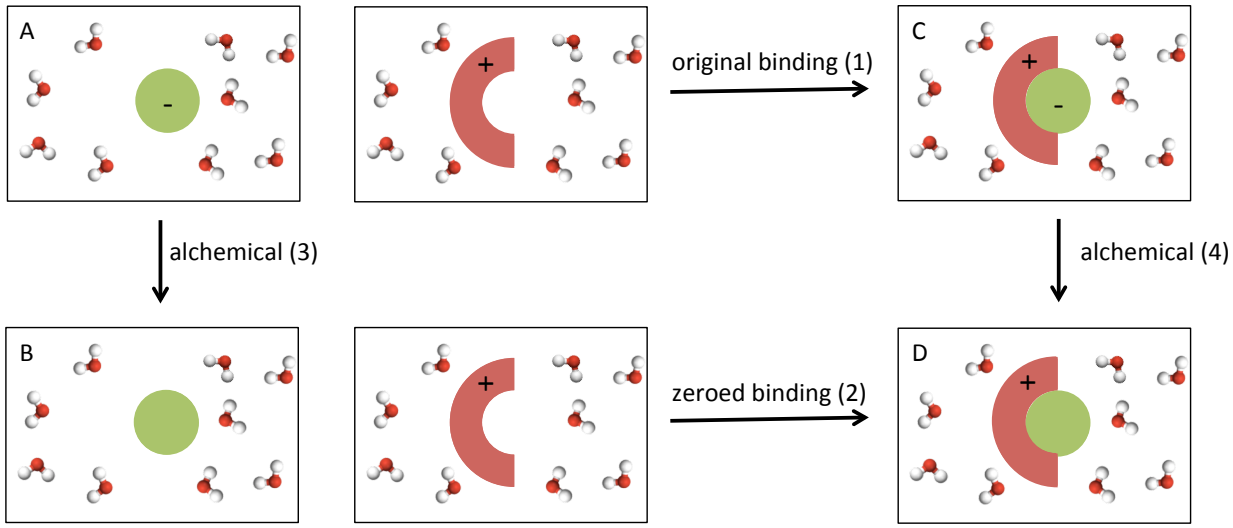


Figure 2: Thermodynamic cycle to calculate the change in free energy, $\Delta\Delta G$, as a residue is charged.

we can get around this issue by carefully selecting a thermodynamic cycle. We can obtain the $\Delta\Delta G$ through the cycle shown in figure 2. The difference in free energy between the horizontal processes in the figure, which corresponds to original formula for $\Delta\Delta G$ in equation (2), is the same as the difference in the vertical processes, as shown in equation (11). Then, we perform the same cycle in the presence of crowders, and use equation (3) to obtain the $\Delta\Delta\Delta G$. In order to perform the alchemical, vertical transformations, where the residue is being charged or uncharged, we must turn to free energy perturbation.

$$\begin{aligned}(2) - (1) &= (D - B) - (C - A) \\ &= D - B - C + A \\ &= (D - C) - (B - A) \\ &= (4) - (3)\end{aligned} \tag{11}$$

Free energy perturbation (FEP) is a statistical mechanical method to compute the free energy difference as one system, A, is transformed into another, B.[43] If a particular physical

process is computationally infeasible to model directly, the free energy of the physical process may be determined by using a thermodynamic cycle in conjunction with alchemical (i.e. nonphysical) transformations. Consider the difference in the free energies of the states, equation (12).

$$\Delta G = G_B - G_A \tag{12}$$

This can be rewritten in terms of the NPT partition function Z, using equation (13), to produce (14). k is the Boltzmann constant and T is the temperature.

$$G = -kT\ln Z \tag{13}$$

$$\Delta G = -kT\ln\frac{Z_B}{Z_A} \tag{14}$$

If the partition functions Z_A and Z_B are known, the free energy difference can be calculated by equation (14). However, the partition functions are often not known. Z_B can be rewritten as shown in equation (15), where r are the degrees of freedom of the system and U is the potential energy, defined to include volume as a coordinate and pressure as a parameter in an NPT ensemble.

$$Z_B = \int e^{-\frac{U_B}{kT}}\,\mathrm{d}r \tag{15}$$

$$= \int e^{-\frac{U_B}{kT}} e^{-\frac{U_A}{kT}} e^{\frac{U_A}{kT}}\,\mathrm{d}r \tag{16}$$

$$= \int e^{-\frac{U_A}{kT}} e^{-\frac{U_B - U_A}{kT}}\,\mathrm{d}r \tag{17}$$

Taking the ratio of partition functions results in equation (18); the $< ... >_A$ notation indicates an average with respect to state A.

$$\frac{Z_B}{Z_A} = \frac{\int e^{-\frac{U_A}{kT}} e^{-\frac{U_B - U_A}{kT}} \mathrm{d}r}{e^{-\frac{U_A}{kT}} \mathrm{d}r} \tag{18}$$

$$= < e^{-\frac{U_B - U_A}{kT}} >_A \tag{19}$$

If equation (14) and equation (19) are combined, the Zwanzig equation (20), also known as exponential averaging (EXP), is produced.

$$\Delta G(A \rightarrow B) = -kT \ln < e^{-\frac{U_B - U_A}{kT}} >_A \tag{20}$$

The EXP method is limited. Notice that, in effect, we are taking the microstates of state A and using them to calculate Z_B. The difference in energy, ΔU, is obtained by averaging over the equilibrium ensemble of the initial state. Thus, we reweigh the states, using $e^{-\frac{U_B}{kT}}$ instead of $e^{-\frac{U_A}{kT}}$. The method converges rapidly only if states A and B have large phase space overlap, meaning that they contain many microstates in common. In that case, the difference in potential energy is small, so the exponential in the Zwanzig equation is large. If they overlap poorly, there is a great deal of statistical uncertainty and the equation converges slowly. In FEP, at each state, a MD simulation is run, to determine the probability of microstates at that state. Because the simulation may not begin at equilibrium, the system must first be minimized and equilibrated at each state. FEP therefore is very computationally intensive. If the perturbation between A and B is large (i.e., the overlap between the two states is small), the simulation must run for a long time to sample the dissimilar microstates. One solution is to select intermediate states; the transformation from state A to B is denoted by the reaction progress coordinate, λ. In our case, λ corresponds to the charge on the residue being zeroed out. When choosing intermediate states for this method, we choose states that maximize overlap and ensure convergence, therefore maximizing the calculations efficiency and accuracy.

The EXP method requires all high probability microstates of A be found in high proba-

bility in B. A less stringent method is the Bennett acceptance ratio method (BAR), which involves sampling from both states A and B instead of just A.[44] (As a result, BAR requires configurational information at both states.) Therefore, BAR is more efficient and requires less phase overlap, though it is less intuitive. The original Bennett paper deals with finding the Helmholtz free energy from a Monte Carlo simulation. Although this derivation is not strictly true for a molecular dynamics simulation, the MD data for a system with many degrees of freedom will be very similar to the configurational distribution.[44]

$$\frac{M(x)}{M(-x)} = e^{-x} \tag{21}$$

For any function M that satisfies equation (21) (i.e., that follows the principle of detailed balance), we can write a version for both U_A and U_B (equations (22) and (23)). After some algebra, we get equation (26) out, which Bennett called the acceptance probability (recall that he was dealing with MC).

$$\frac{M(U_A)}{M(-U_A)} = e^{-U_A} \tag{22}$$

$$\frac{M(U_B)}{M(-U_B)} = e^{-U_B} \tag{23}$$

$$\frac{M(U_A)}{M(-U_A)}e^{-U_B} = \frac{M(U_B)}{M(-U_B)}e^{-U_A} \tag{24}$$

$$M(U_A)M(-U_B)e^{-U_B} = M(U_B)M(-U_A)e^{-U_A} \tag{25}$$

$$M(U_A - U_B)e^{-U_B} = M(U_B - U_A)e^{-U_A} \tag{26}$$

If the relation in equation (26) is integrated over all of configuration space, and then multiplied by $\frac{Z}{Z}$, we get (27).

$$Z_A \frac{\int M(U_A - U_B)e^{-U_B} dq_1 ... dq_2}{Z_A} = Z_B \frac{\int M(U_B - U_A)e^{-U_A} dq_1 ... dq_2}{Z_B} \tag{27}$$

Notice that the fractional portions of equation (27) are averages; this leads to the acceptance probability ratio equation, equation (28).

$$\frac{Z_A}{Z_B} = \frac{< M(U_A - U_B) >_B}{< M(U_B - U_A) >_A} \tag{28}$$

Now, a more general version of equation (28) can be obtained by allowing an arbitrary weighting function $W(q_1...q_N)$ to be included in both numerator and denominator. This results in equation (29).

$$\frac{Z_A}{Z_B} = \frac{Z_A \int W e^{-U_A - U_B} \mathrm{d}q^N}{Z_B \int W e^{-U_B - U_A} \mathrm{d}q^N} \tag{29}$$

$$= \frac{< M(U_A - U_B) >_B}{< M(U_B - U_A) >_A} \frac{\int W e^{-U_A - U_B} \mathrm{d}q^N}{\int W e^{-U_B - U_A} \mathrm{d}q^N} \tag{30}$$

$$= \frac{< W e^{-U_A} >_B}{< W e^{-U_B} >_A} \tag{31}$$

Bennett's paper[44] then finds the W which minimizes the error in the estimated free energy. The rest of the derivation is briefly outlined and presented without proof; for more details, consult Bennett's paper. The free energy is shown in equation (14). If the data are n_A configurations in state A and n_B configurations in state B, and the sample size is large, then the error will be approximately Gaussian and expected to look like equation (32). Equation (34) shows the result of optimizing W; if the result is substituted back into equation (31), then equation (35) follows.

$$(\frac{\Delta G_{est} - \Delta G}{kT})^2 = \frac{< W^2 e^{-2U_B} >_A}{n_A (< W e^{-U_B} >_A)^2} + \frac{< W^2 e^{-2U_A} >_B}{n_B (< W e^{-U_A} >_B)^2} - \frac{1}{n_A} - \frac{1}{n_B} \tag{32}$$

$$= \frac{\int (\frac{Z_A}{n_A} e^{-U_B} + \frac{Z_B}{n_B} e^{-U_A}) W^2 e^{-U_A - U_B} \mathrm{d}q^N}{(\int W e^{-U_A - U_B} \mathrm{d}q^N)^2} - \frac{1}{n_A} - \frac{1}{n_B} \tag{33}$$

$$W = \frac{\text{constant}}{\frac{Z_A}{n_A}e^{-U_B} + \frac{Z_B}{n_B}e^{-U_A}} \tag{34}$$

$$\frac{Z_A}{Z_B} = \frac{< \frac{\text{constant}}{\frac{Z_A}{n_A}e^{-U_B}+\frac{Z_B}{n_B}e^{-U_A}} e^{-U_A} >_B}{< \frac{\text{constant}}{\frac{Z_A}{n_A}e^{-U_B}+\frac{Z_B}{n_B}e^{-U_A}} e^{-U_B} >_A} \tag{35}$$

$$= \frac{< \frac{1}{1+(e^{U_A-U_B})(\frac{Z_A n_B}{Z_B n_A})} >_B}{< \frac{1}{1+(e^{U_B-U_A})(\frac{-Z_A n_B}{Z_B n_A})} >_A} \frac{Z_A n_B}{Z_B n_A} \tag{36}$$

$$= \frac{< f(U_A - U_B + C) >_B}{< f(U_B - U_A - C) >_A} e^C \tag{37}$$

Equation (37) uses the Fermi function $f(x) = \frac{1}{1+e^x}$ and a shift constant $C = \ln\frac{Z_A n_B}{Z_B n_A}$. The shift constant C must be determined self consistently. Finally, the free energy can be determined by equation (38).

$$\Delta G = kT(\ln\frac{< f(U_A - U_B + C) >_B}{< f(U_B - U_A - C) >_A} + C) \tag{38}$$

1.5 Comparison to implicit solvent

The ultimate goal of this work is to perform component analysis of barstar residues in the barnase-barstar system, in explicit solvent, in the presence and absence of crowders. We can then assess how the water depletion and loss of water mobility caused by the crowders effects the contribution of the charges on particular residues to the binding free energy. In order to perform component analysis in explicit solvent, free energy perturbation is necessary to calculate the free energy change. When taken in context of previous work, which dealt with water depletion, we will be able to see if the loss of water mobility in the presence of crowders changes this contribution.

Implicit solvent calculations were largely done with the same methodology as in our previous study.[30] In order to compare the implicit and explicit solvent results in a controlled way,

we changed some parameters for the implicit solvent calculations, as described in Methods. The results of these calculations are reported in the current study.

At this point, component analysis has been performed for the aspartic acid 35 residue of barstar in the absence of crowders.

2 Methods

2.1 Structure

The structure used in this study was modified from a 2.0 Å resolution crystal structure of the barnase-barstar complex, containing a Cys to Ala (40,82) double mutant of barstar (PDB ID 1BRS).[39] The original crystal structure contains three barnase-barstar complexes; chains A and D, which have the least missing density, were used. Crystallographic water was removed for both explicit and implicit solvent calculations. Hydrogens were modeled onto the structure with the HBUILD[45] function of CHARMM,[46] using the CHARMM22 force field[42] and the TIP3P water model.[47] The two N-terminal residues of barnase and residues 64 and 65 of barstar were not resolved in the crystallographic structure and were built in using MODELLER.[48] Missing side chain density was added via CHARMM.

For static, implicit solvent calculations and as the starting structures for the MD simulation, the following changes were made to the structure. Based on potential hydrogen bonding, the amide groups of asparagine and glutamine residues and the imidazole groups of histidine residues were flipped if necessary, and the tautomerization state of histidines were assigned. Protonated states were chosen based on physiological pH: lysine and arginine side chains were protonated, while aspartic acid and glutamic acid were not. Charged termini were used. Note that the resulting structure differs from the structure used in our previous study[30] because missing residues, including missing terminal residues, were built back in in this study.

A molecular dynamics simulation was run on the resulting structure with GROMACS.[49,50]

The Gromos96 43a1 force field[51] was used. The structure was placed in a cubic box, where the sides of the box were at least 0.8 nm away from the complex, with a 0.1 M concentration of salt ions (sodium and chloride) in water. It was minimized for 100 steps using a steepest descents minimization.

The resulting structure ("structure with unaltered interface") was used to perform FEP. An attempt was made to run an initial MD simulation on this structure prior to the FEP, but the resulting structure had an altered interface ("structure with altered interface"). Because the structure with altered interface was also used to perform FEP and some of those results are discussed, I will state the parameters used in this initial MD simulation to produce the structure with altered interface. An MD simulation with the following parameters was run on the minimized structure ("structure with unaltered interface") for 20 ns with a step size of 0.002 ps. Non-bonded interactions were calculated every 10 steps, using the grid algorithm (i.e., only neighboring grid cells are used when constructing the new neighbor list). The cut-off distance for Coulombic and van der Waals interactions was 1 nm (note that when using PME, the cut-off distance for Coulombic interaction indicates where real and Fourier space are used). Periodic boundary conditions were used in the x, y, and z directions. The temperature was kept at 310 K using a Berendsen thermostat[41] with a coupling time of 0.1 ps. The Berendsen thermostat re-scales the velocities of the particles in the simulation, so the system behaves as if it were weakly coupled to a heat bath at the preset temperature. The pressure was coupled to a barostat at 1 bar with a coupling time of 1 ps. All bonds were constrained with the LINCS algorithm.[37]

2.2 Free Energy Perturbation

Referring back to the overall thermodynamic cycle, figure 2, in order to determine the $\Delta\Delta\Delta G$ for a particular residue, four perturbations from a charged version of the residue to an uncharged one are necessary, in the following scenarios:

1. Barnase-barstar complex, no crowders

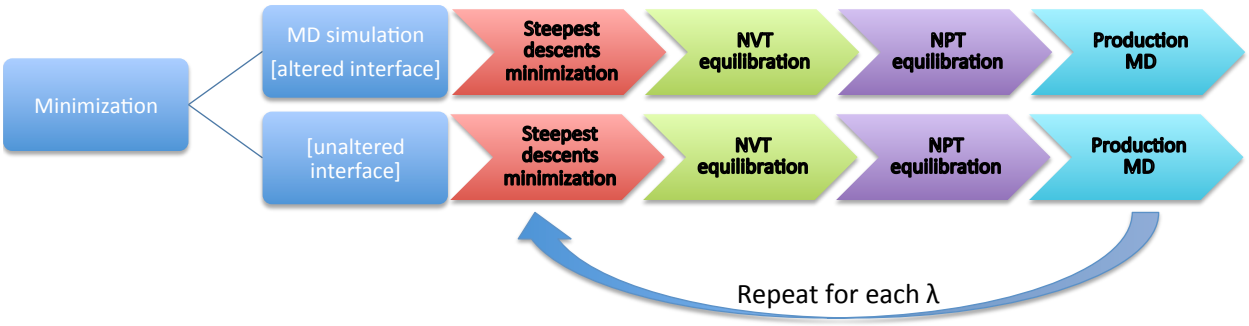


Figure 3: Simulations run for each λ value.

2. Barstar only, no crowders
3. Barnase-barstar complex, crowders
4. Barstar only, crowders

In this thesis, we describe efforts to perform perturbations in scenarios 1 and 2 using barstar's ASP35 residue. Recall that the reaction coordinate λ corresponds to the process of perturbing the system from the charged ASP35 to the uncharged ASP35. Since aspartic acid has a -1 overall charge, a counter ion within solvent was used, which had the appropriate charge to keep the system neutral, starting out with a +1 overall charge at $\lambda = 0$ and ending with a neutral overall charge at $\lambda = 1$. The counter ion had the same Lennard Jones potential and size as a sodium ion. Because we ultimately want to compare the results in explicit solvent to those in the implicit solvent, where rigid binding was assumed,[30] we did not want to allow the proteins to move about. During the course of the perturbation, the positions of the atoms in the barnase-barstar complex were restrained to their starting positions, using a harmonic oscillator potential with a force constant of 1000 $\frac{\text{kJ}}{\text{mol nm}^2}$.

Before the production MD at each λ, the system was first minimized and equilibrated in the following steps, also shown in figure 3. This workflow is based on Justin Lemkul's GROMACS tutorial.[52] In all cases, periodic boundary conditions were used in the x, y, and z directions.

First, a 5000 step steepest descents minimization was run with a maximum step size

of 0.01 $\frac{\text{kJ}}{\text{mol nm}}$ and a tolerance of 100 $\frac{\text{kJ}}{\text{mol nm}}$. The electrostatics were calculated using the particle mesh Ewald method with a distance cutoff of 1 nm and an interpolation order of 6. The van der Waals force was calculated using a normal LJ potential until 0.8 nm, after which it is smoothly switched off to zero at 0.9 nm. A long range dispersion correction was applied to both energy and pressure. Both temperature and pressure coupling were off, as is appropriate for energy minimizations. Bonds were constrained with the LINCS algorithm.[37] This energy minimization "relaxed" the structure to prevent steric clashes, inappropriate geometry, and other major issues. The result was a reasonable structure with respect to geometry.

Second, an NVT equilibration was run for 200 ps with a time step of 2 fs. The electrostatic and van der Waals forces were calculated as described in the steepest descents minimization. The leap-frog stochastic dynamics integrator was used, which sets the temperature of the system at 300 K. Pressure coupling was turned off. The equilibration was necessary because the reasonable structure produced by the energy minimization may not have given the solvent enough time to reorient around the protein, crowders, and ions. Additionally, the temperature of the system, as calculated from kinetic energies, was not yet correct. An NVT equilibration will in general bring and stabilize the system to a desired temperature.

Finally, an NPT equilibration was run for 200 ps with a time step of 2 fs. All settings were the same as the NVT equilibration, but with a pressure coupling using the Berendsen method, which involves exponential relaxation pressure coupling with time constant 0.5 ps. This was necessary to stabilize the pressure and density of the system. The result after the energy minimizations and equilibrations is a system that is equilibrated to a desired temperature and pressure.

The production MD was run using the leap-frog stochastic dynamics integrator for 5 ns with a time step of 2 ps, and the same settings as the NPT equilibration.

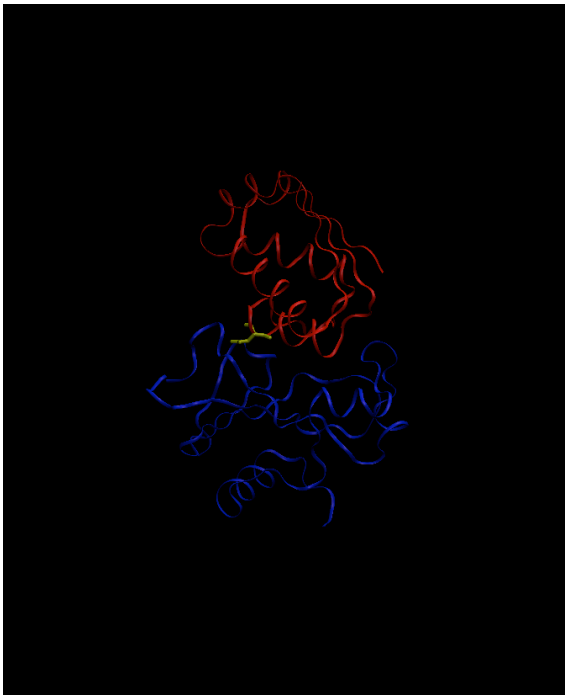
Because the effect of discharging the residue is greatest when its charge is larger, an effort was made to sample more finely at low λ values, which corresponds to greater magnitude

charges. Due to time constraints, not all the desired trials have been run; future work is discussed later. For the structure with altered interface, 24 λ values were used: 21 equally spaced between 0 and 1 at intervals of 0.05, and additional λs of 0.025, 0.075, and 0.125. For the structure with unaltered interface, 21 λ values were used, equally spaced between 0 and 1 at intervals of 0.05.

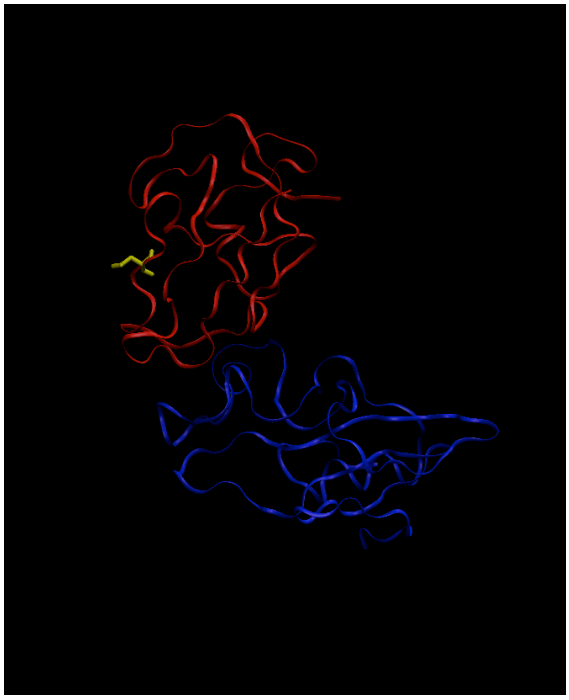
The FEP itself was done using the Bennett's acceptance ratio method (BAR).[44]

2.3 Implicit solvent calculations

In order to assess the effect of water mobility, we needed to compare explicit solvent results to implicit solvent results; in order to compare results in a controlled way, we need to modify the parameters used in our previous study.[30] Implicit solvent calculations were done with the same methodology as in our previous study,[30] with a few changes to make the system consistent with the explicit solvent parameters. A salt concentration of 0.1 M was used, because the explicit solvent simulations were done at 0.1 M salt concentration. A dielectric constant of 1 was used for the protein, instead of the 4 used in the previous study. This is because the dielectric constant of 4 was supposed to account for the polarizability and motion of the protein; however, in these explicit solvent simulations, we are keeping the protein restrained. Because of memory limitations, a 401x401x401 grid was used when solving for the potentials using the finite-difference method. During the component analysis, the entire residue's charges were zeroed, as opposed to only the side chain, in both explicit and implicit calculations. Additionally, the structure used for the implicit solvent calculations here was slightly different than the one described in Qi *et al.* it had additional residues built in and was minimized, as described in the Methods section. As a result, all crystallographic waters were removed.



(a) Barnase-barstar after minimization, with unaltered interface.



(b) Barnase-barstar after a 20 ns MD simulation, with altered interface.

Figure 4: Barnase-barstar structures used for the FEP. Barstar is shown in red; barnase in blue. Barstar's ASP35 is shown in yellow.

2.4 Analysis, visualization, and computational details

All visualizations of molecules were done using VMD.[53] Graphs were generated with MATLAB.[54]

All simulations were run with 16 3000 MHz AMD Opteron processors.

3 Results and Discussion

When an unrestrained MD simulation was run on the barnase-barstar complex using the parameters described in section 2.1, the interface changes, as seen in figure 4b. Notice that the ASP35 that was once in the interface moved away from the interface. In a previous study, using implicit water,[30] we found that the aspartic acid had a large $\Delta\Delta G$, as it was in the interface and therefore very desolvated upon binding. Also, once the aspartic acid

moved away from the interface, it was no longer be able to interact with barnase as strongly. Nevertheless, because it was the initial FEP analysis run for this thesis, I will discuss the analyses run to confirm that the FEP worked using this structure (altered interface), in addition to discussing the results from a structure more consistent with the model complex used in the implicit solvent study (unaltered interface).

3.1 Implicit solvent

The implicit solvent data for the ΔG of binding between barnase and barstar is presented in kJ/mol. The zeroed residue was barstar's ASP35. An in depth explanation and analysis of similar trends in the ligand desolvation penalty (LDP), receptor desolvation penalty (RDP), and interaction (INT) may be found in Qi *et al.*[30]

	LDP	RDP	INT	TOT
$\Delta G_{original}$	535.8	529.7	-747.1	318.4
ΔG_{zeroed}	356.4	529.7	-540.7	345.4
$\Delta\Delta G$	-179.4	0	206.4	26.97

The implicit solvent $\Delta\Delta G$ of 26.97 kJ/mol is different than previously reported $\Delta\Delta G$ values for this residue[30,32,55] because the parameters used were slightly different – the inner dielectric was 1, salt concentration was 0.1 M, and the entire residue's charge was zeroed out. As a result, the desolvation penalties and interactions were magnified, but the total $\Delta\Delta G$ decreased. In any case, note that the sign of $\Delta\Delta G$ is positive, indicating that the charges of the aspartic acid contributed favorably to the binding of barnase and barstar.

3.2 λ=0 simulation analysis, altered interface structure

The following analyses were done for the $\lambda = 0$ value.

Figure 5 shows that during the energy minimization, the potential energy converged. During the production MD, we should see the temperature, pressure, and density remain

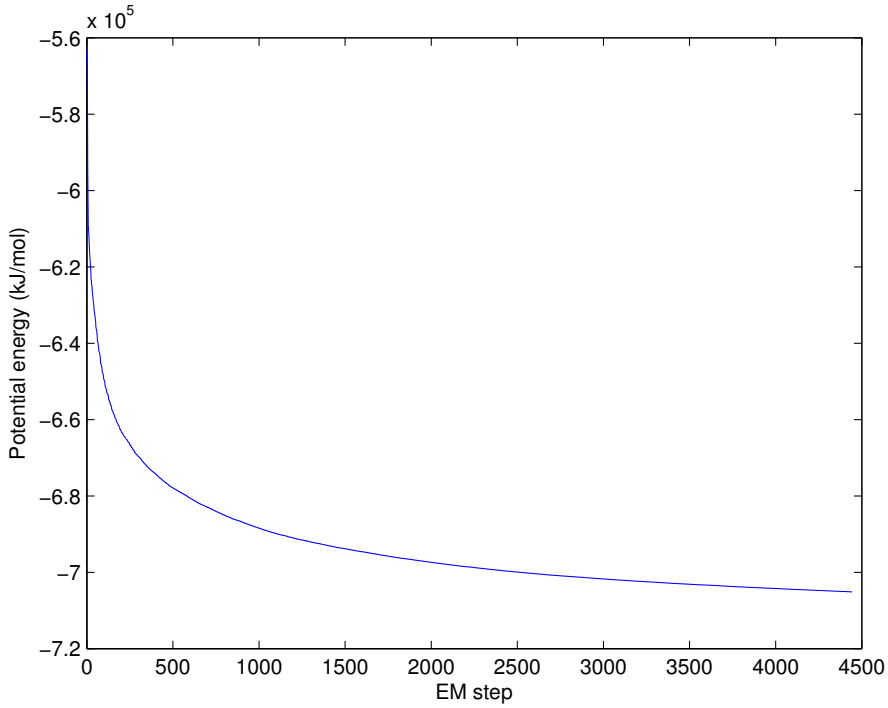


Figure 5: Potential energies during the energy minimization for λ=0.

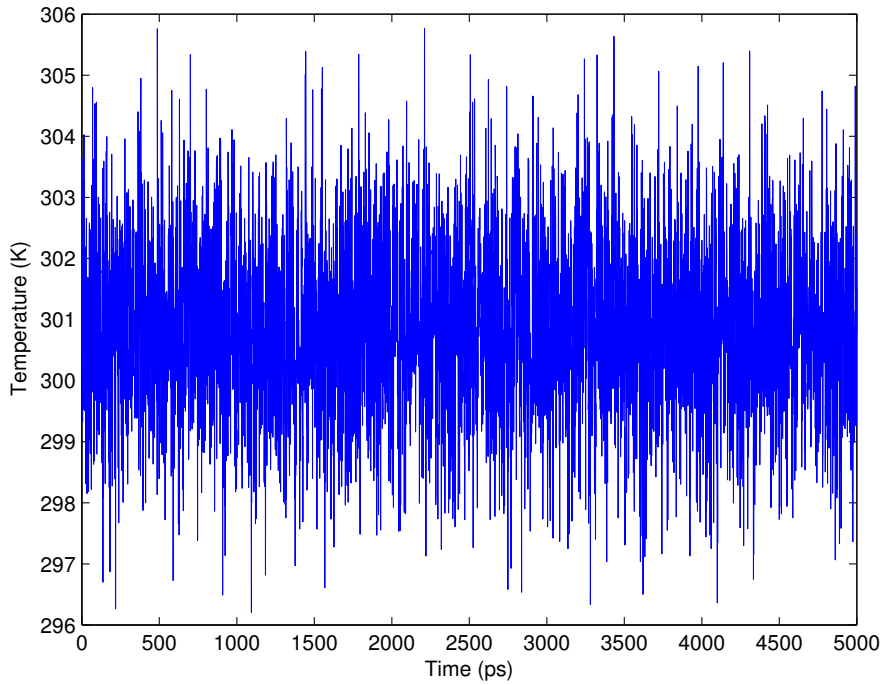


Figure 6: Temperature during the production MD for λ=0.

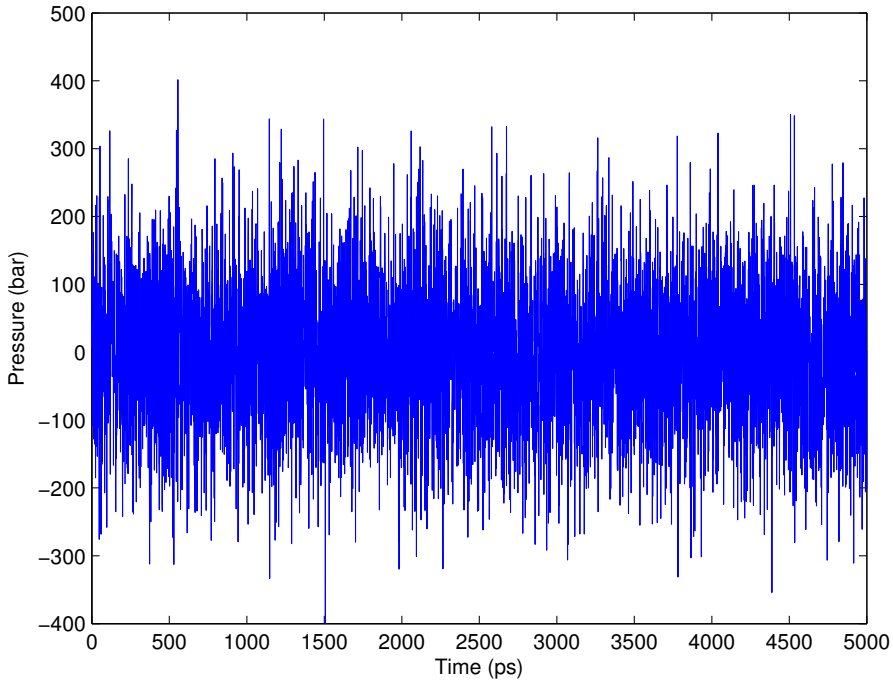


Figure 7: Pressure during the production MD for λ=0.

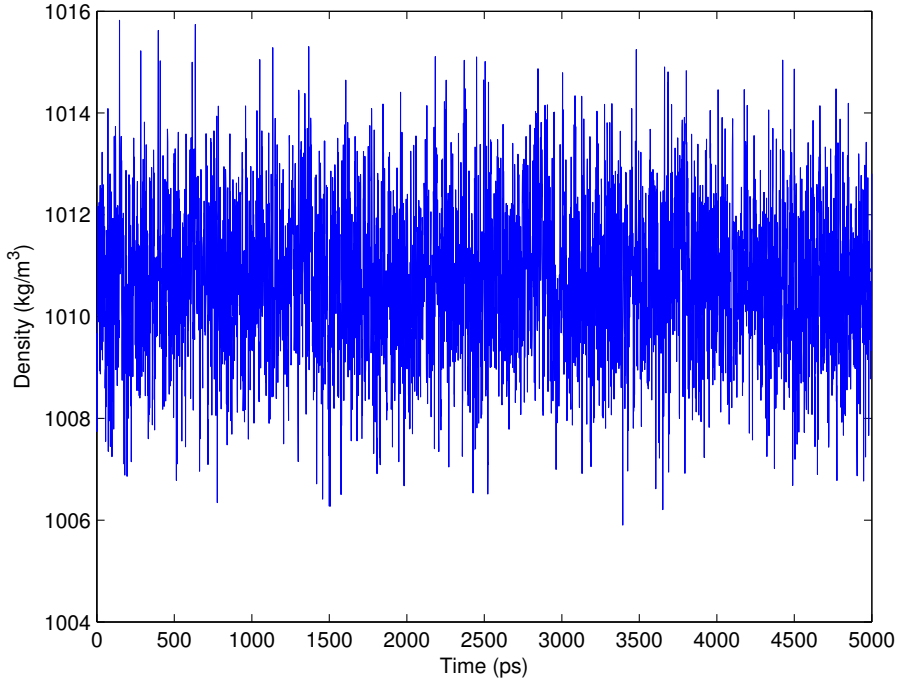


Figure 8: Density during the production MD for λ=0.

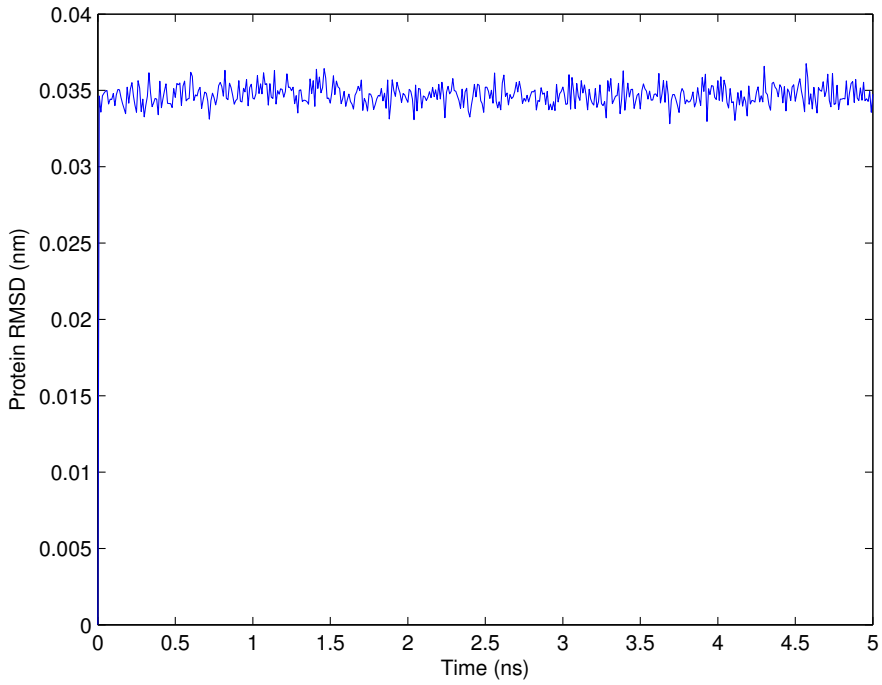


Figure 9: RMSD of the protein during the production MD for λ=0.

about the same on average, though they may fluctuate. Figure 6 shows the temperature fluctuated around 300.9 K, which was approximately the target temperature, 300 K. The pressure (figure 7) fluctuated wildly and had an average at -1.9 bar. These large pressure fluctuations are expected because it takes a large pressure change to even slightly change the volume of a liquid, and equilibrium is not instantaneous. Small volume changes therefore cause large pressure changes. Additionally, instantaneous pressure is not well defined, especially over a short (ps) time scale; pressure is only meaningful as a macroscopic quantity and as a time average. Fluctuations on the order of hundreds of bar, like we observed, are therefore typical. The negative pressure indicates that the density of the system was too low and the system wanted to contract; it is not unusual. Because of these known properties of the pressure during an NPT simulation, the density is typically considered more useful. The density of water by itself should be 1000 $\frac{\mathrm{kg}}{\mathrm{m}^3}$. The average density, seen in figure 8, was 1011 $\frac{\mathrm{kg}}{\mathrm{m}^3}$.

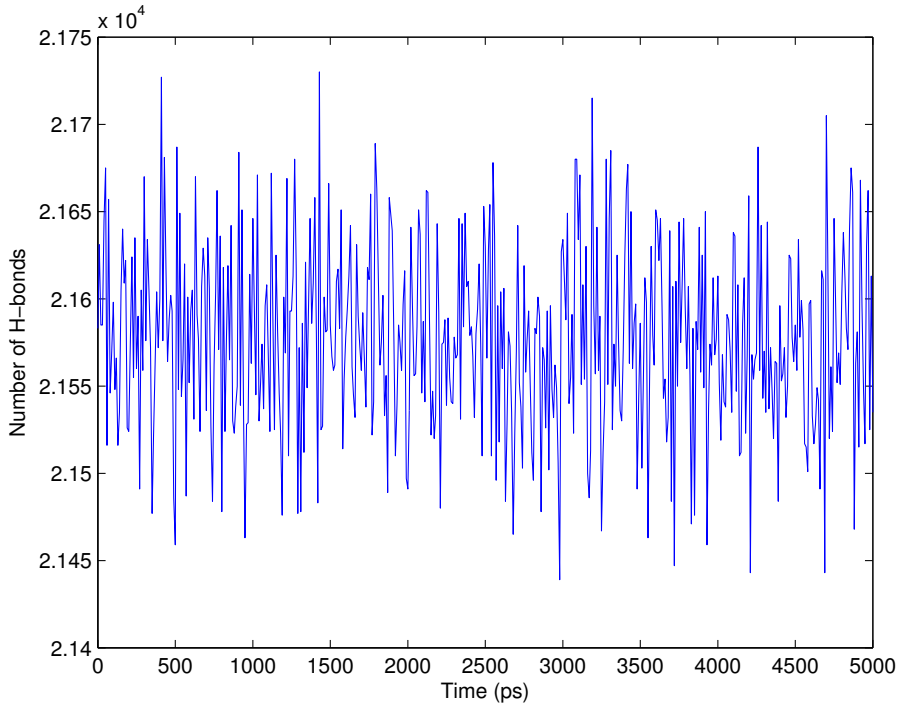


Figure 10: Number of hydrogen bonds between solvent molecules during the production MD for λ=0.

In figure 9, the RMSD of the protein during the production MD suggests that the structure was stable. Because the protein atom positions were restrained and not constrained, the atoms of the protein did move around slightly. However, RMSD will not tell us if the solvent had equilibrated properly, as we would expect the water and ions to move about freely. We would instead see the RMSD of the solvent slowly reach some maximum no matter what structure the RMSD was calculated relative to. Instead, the density, pressure, and the number of H-bonds (figure 10) being relatively constant indicates solvent equilibrium.

3.3 FEP analysis, altered interface structure

The following results are again for the barnase-barstar structure with altered interface, during the discharging of the ASP35 residue of barstar.

In figure 11, the ΔG between each λ value is plotted. As expected, the ΔG was greatest at low values of λ, and sampling more there was appropriate. (Note that because the error

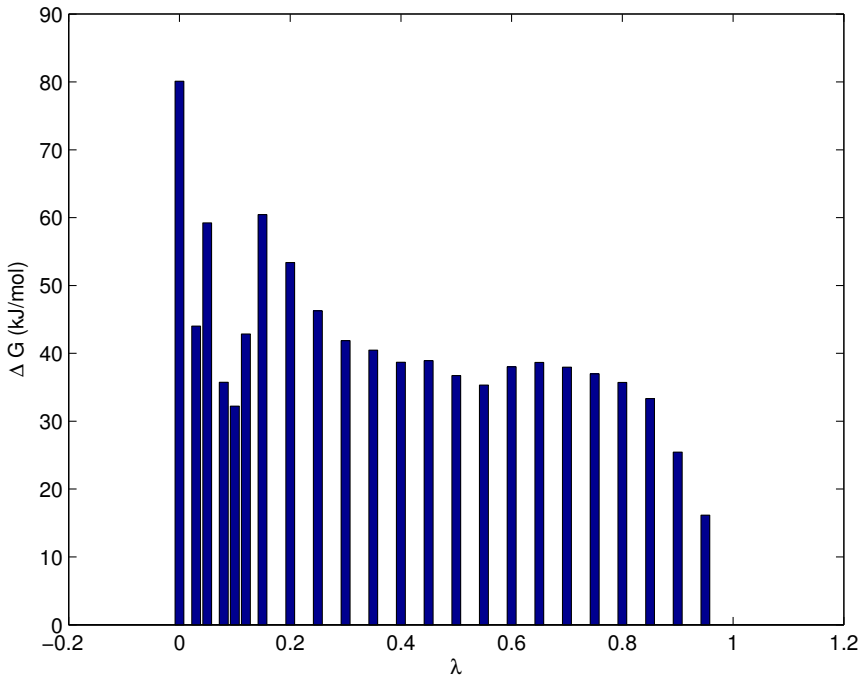


Figure 11: Free energy difference between neighboring values of λ for the bound barnase-barstar complex. The sum of these ΔGs corresponds to the $\Delta G_{alchemical,barnasebarstar}$.

associated with the phase space overlap is always less than 2 kJ/mol, the error bars are not visible.)

In figure 12, the distributions of $\frac{\delta H}{\delta \lambda}$ for each value of λ are plotted. (Note that here, H is the Hamiltonian, not enthalpy.) From these $\frac{\delta H}{\delta \lambda}$'s, the ΔH's can be calculated between neighboring λ values. In figure 13, the distributions of ΔH values is graphed. Notice that the ΔHs for a particular λ are calculated from configurations in a different λ trajectory; this is denoted by the legend entry. A legend entry of "N($\Delta H(\lambda = 0.5)|\lambda = 0.45$)" means that the ΔH distribution is calculated at a λ value of 0.5 from configurations in the λ=0.45 trajectory. Adequate sampling within each window (indicated by the $\frac{\delta H}{\delta \lambda}$ distribution) and between neighboring windows (indicated by the ΔH distribution) is necessary for a good estimate of ΔG.

In the graphs of ΔH and $\frac{\delta H}{\delta \lambda}$ distributions, notice that especially at low λ, the overlap between neighboring distributions is not very good. However, the error in the $\Delta G_{alchemical}$

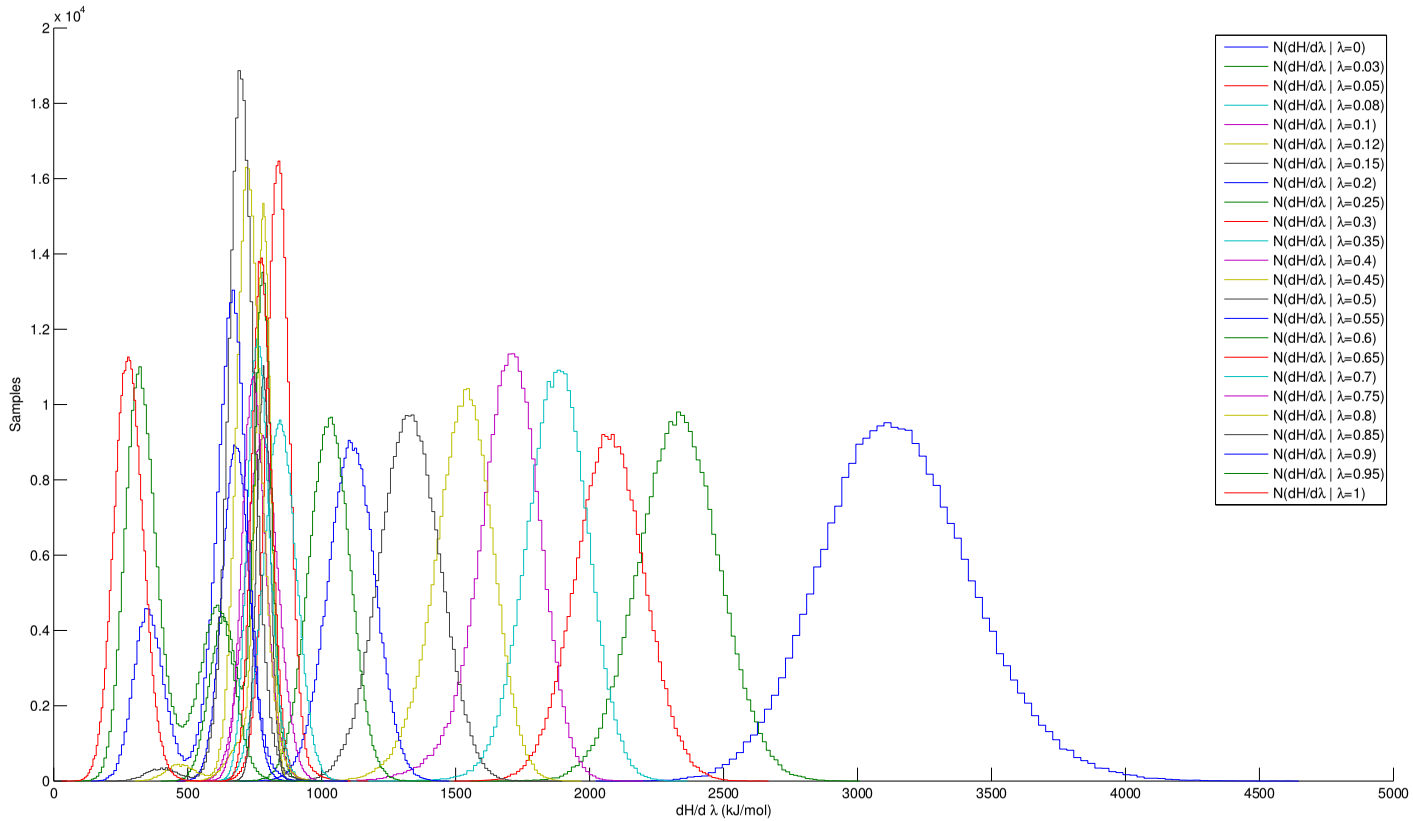


Figure 12: $\frac{\delta H}{\delta \lambda}$ values, colored for the different λs

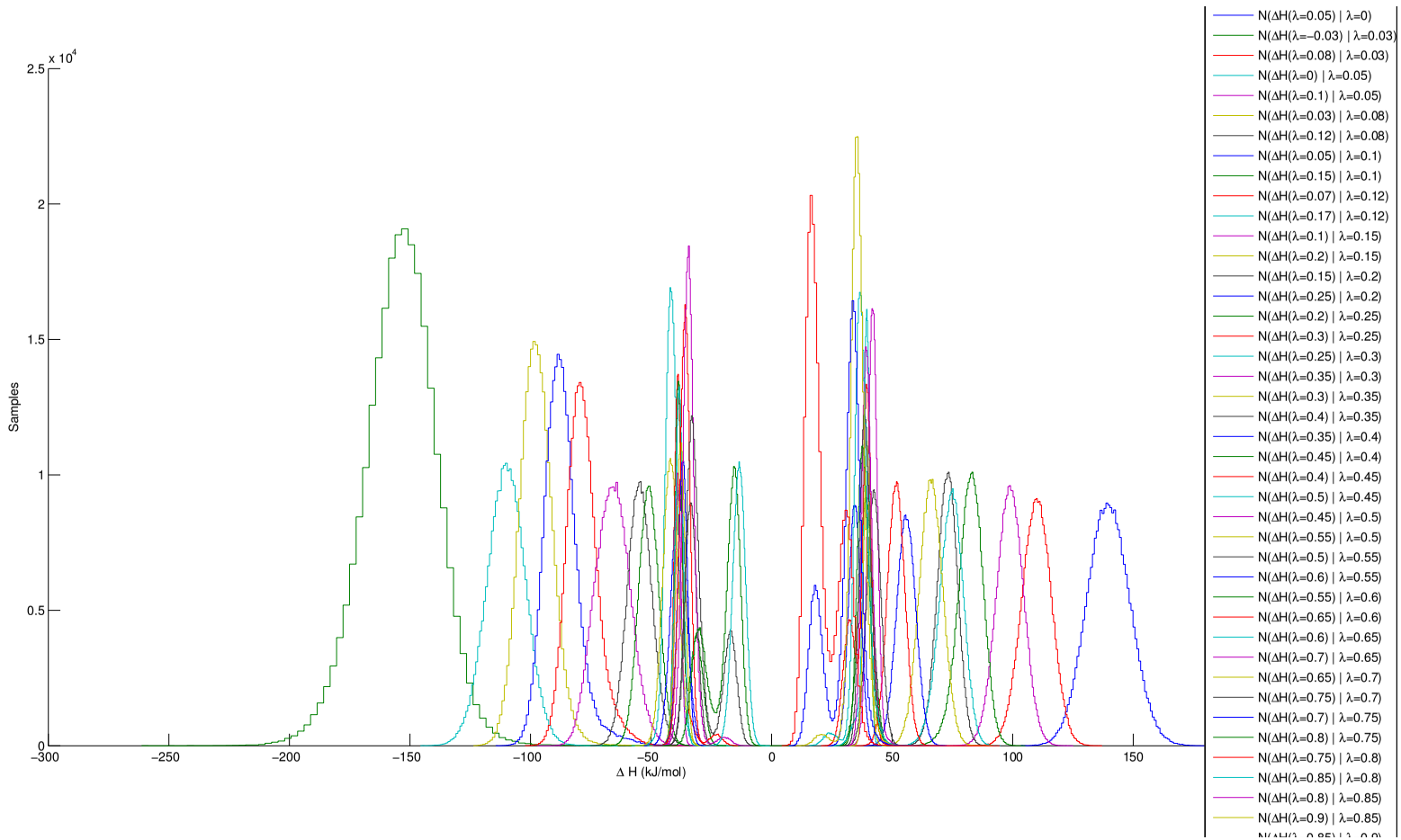


Figure 13: ΔH values, colored for the different λs. Because of the large number of data sets, the legend is cut off

values were only on the order of a few kJ/mol; this error is solely the one associated with the phase space overlap. The derivation for how to arrive at this error is shown in Bennett's paper.[44] Do note, however, that this error assumes uncorrelated data. In the structure with altered interface, going from the 21 equally spaced λ values to 24 λ's (adding three more λ runs at 0.025, 0.075, and 0.125), the $\Delta G_{alchemical}$ goes from 947 kJ/mol to 949 kJ/mol. Part of the reason there is relatively low error is because BAR is effective even at lower phase overlaps.

$\Delta G_{alchemical,barnasebarstar}$	949 ± 4 kJ/mol
$\Delta G_{alchemical,barstar}$	959 ± 4 kJ/mol
$\Delta\Delta G$	-10. ± 9 kJ/mol

The $\Delta\Delta G$ for the structure with altered interface is a small negative value. The small value is expected, because the ASP35 residue has moved away from the interface and is no longer interacting as much with barnase. However, the sign is unexpected – it seems to indicate that the the residue's charges contribute unfavorably to binding. This discrepancy will be discussed further in the unaltered interface results.

3.4 Unaltered interface results

Figures 14, 15, 16, 17, and 18 correspond to the same analysis as in section 3.2 for the barnase-barstar structure at $\lambda = 0$ value, but with the structure that had an unaltered interface (figure 4a). The ΔG between each λ value is shown in figure 21.

Similar figures may be generated for the FEP done on the barstar by itself; they show the same types of fluctuations and trends.

$\Delta G_{alchemical,barnasebarstar}$	920 ± 3 kJ/mol
$\Delta G_{alchemical,barstar}$	954 ± 5 kJ/mol
$\Delta\Delta G$	-34 ± 8 kJ/mol

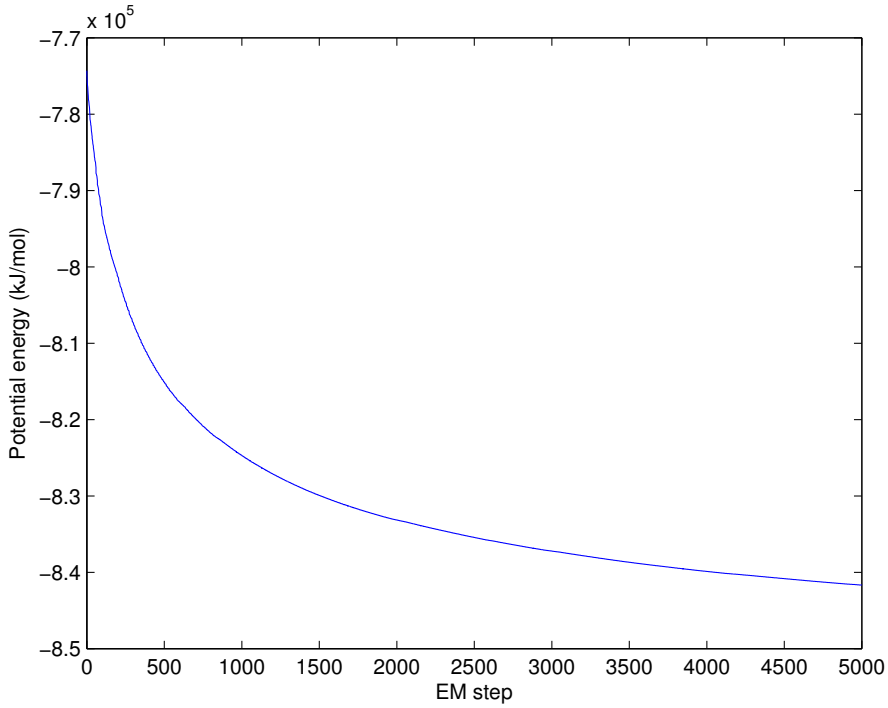


Figure 14: Potential energies during the energy minimization for λ=0.

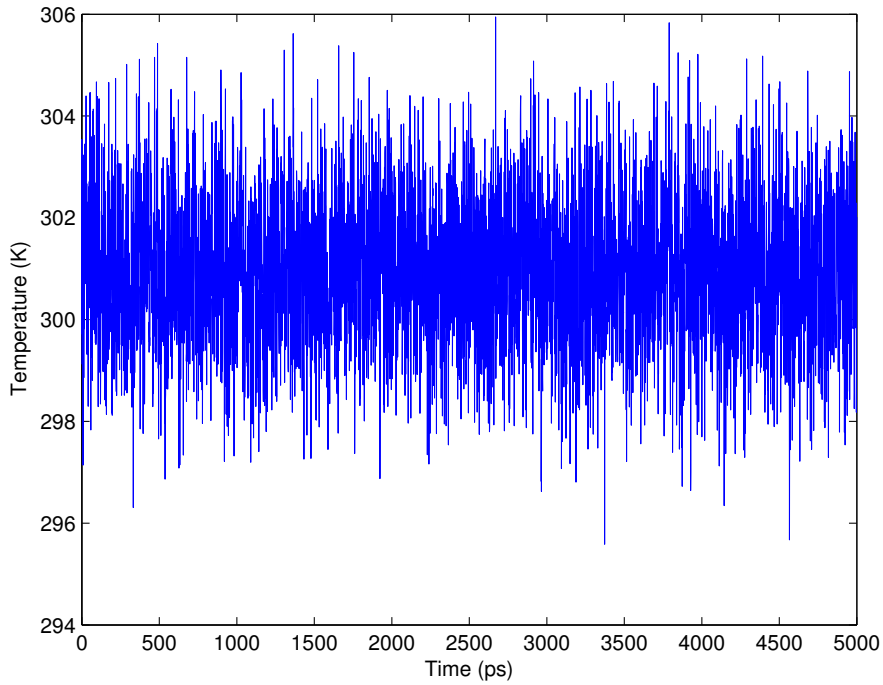


Figure 15: Temperature during the production MD for λ=0. Average temperature is 301 K.

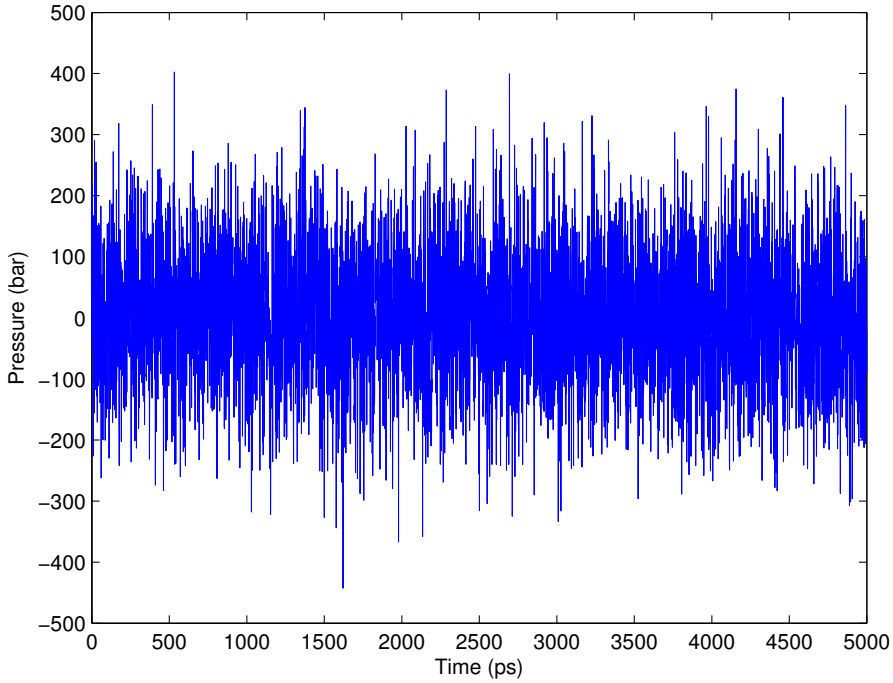


Figure 16: Pressure during the production MD for λ=0. Average pressure is 0.995 bar.

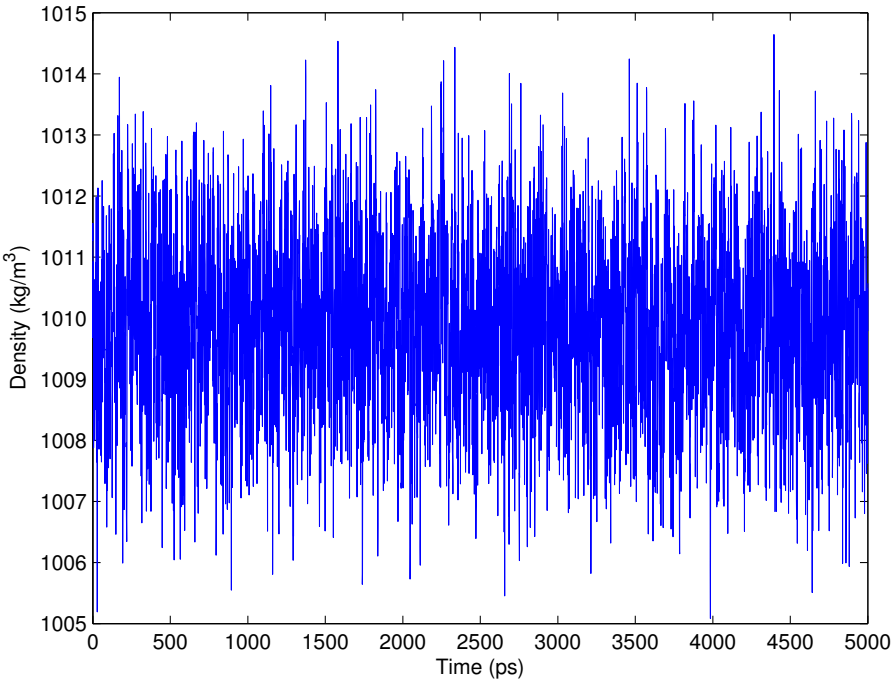


Figure 17: Density during the production MD for λ=0. Average density is 1010 $\frac{\mathrm{kg}}{\mathrm{m}^3}$.

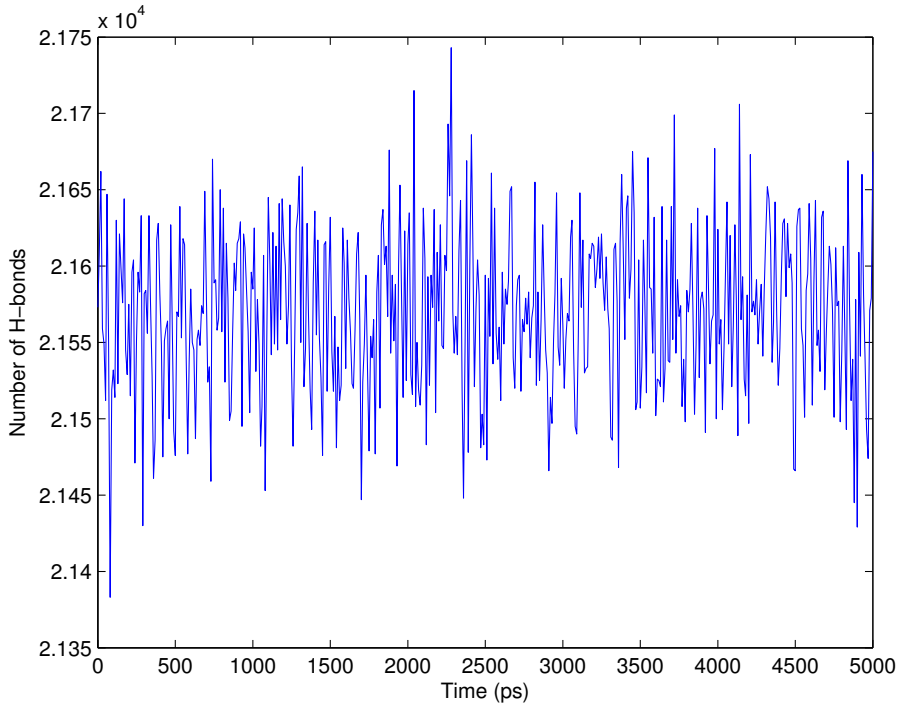


Figure 18: Number of hydrogen bonds between solvent molecules during the production MD for λ=0.

We find that the the charge on the aspartic acid has a smaller effect on the binding in the structure with altered interface than in the structure with the unaltered interface. This is consistent with our expectations: the aspartic acid is now no longer interacting with the barnase, so it should have a smaller impact on binding. However, what is unexpected is that the sign of $\Delta\Delta G$ is negative, which seems to indicate that the aspartic acid's charges are working against binding. This disagrees with our previous implicit solvent study[30] and with previous literature values[32] and so we are actively trying to determine the cause of this result.

When we examine the $\Delta\Delta G$ values between each λ, shown in figure 22, the most negative contributions to the $\Delta\Delta G$ are at low λ values.

One potential problem is the equilibration; because of the way the unbound barstar system was derived from the barnase-barstar complex, the waters were not as well equilibrated in the starting complex. At the various λ values, in both the barnase-barstar complex and

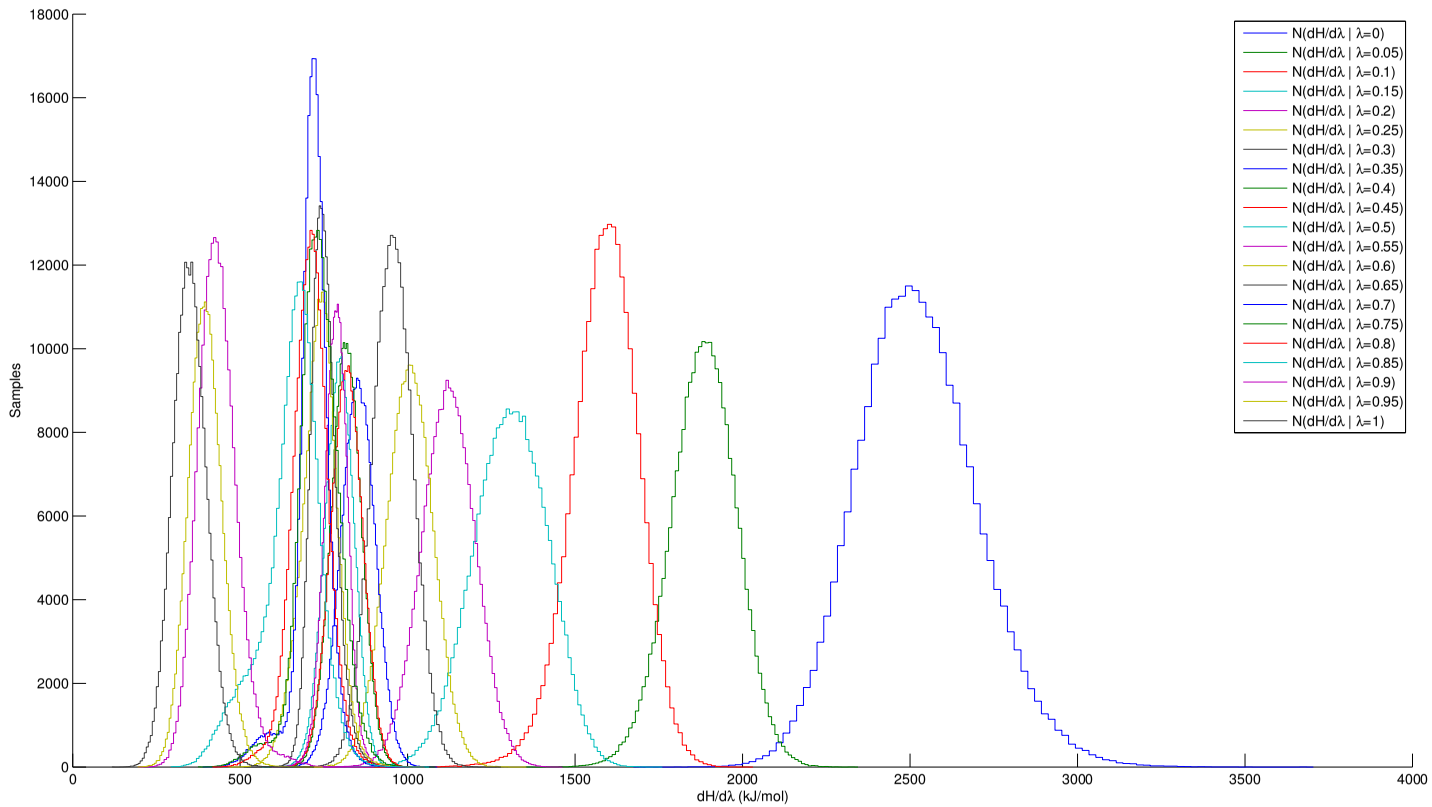


Figure 19: $\frac{\delta H}{\delta \lambda}$ values, colored for the different λs.

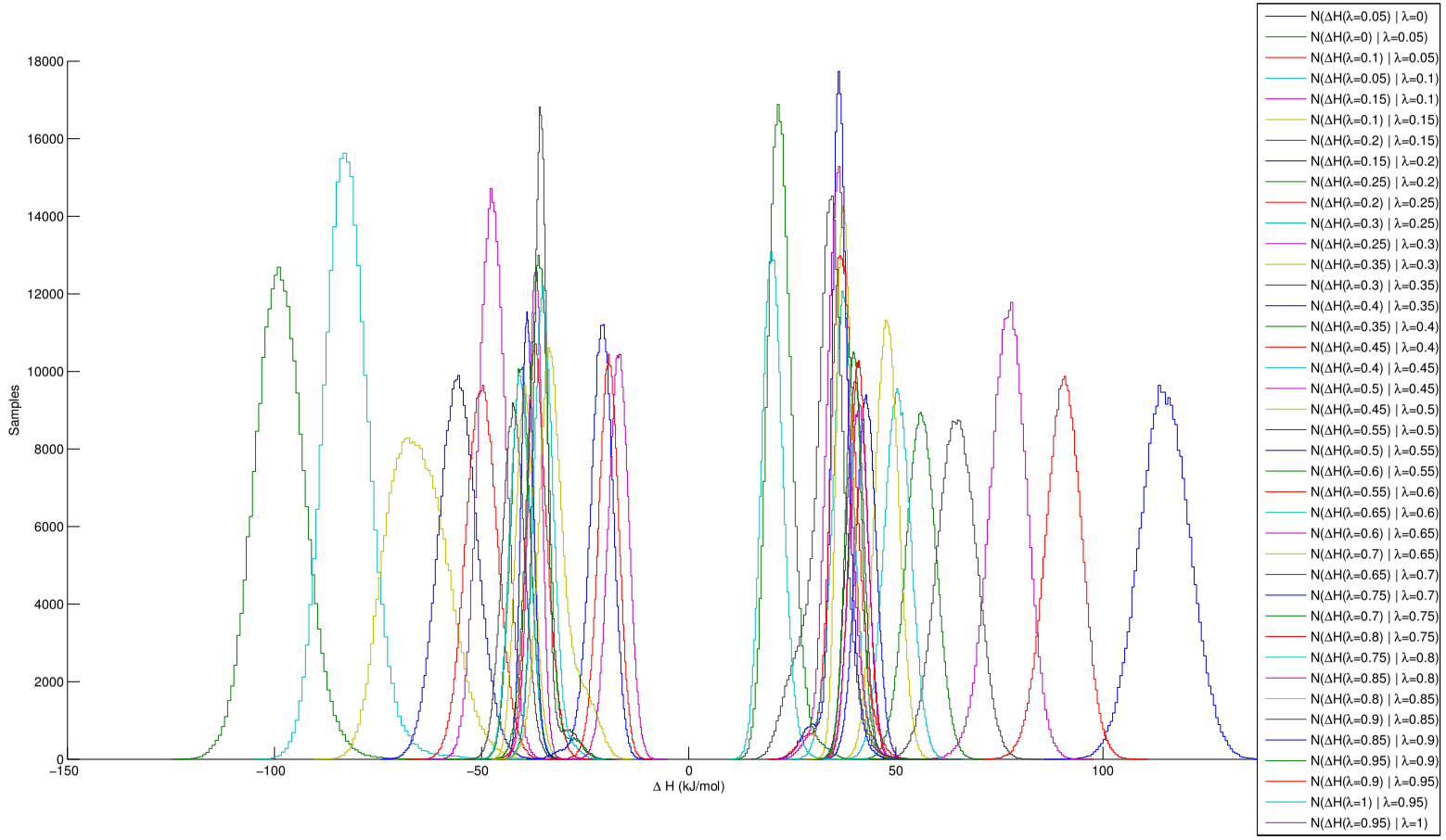


Figure 20: ΔH values, colored for the different λs. Notice that the ΔHs for a particular λ are calculated from configurations in a different λ trajectory.

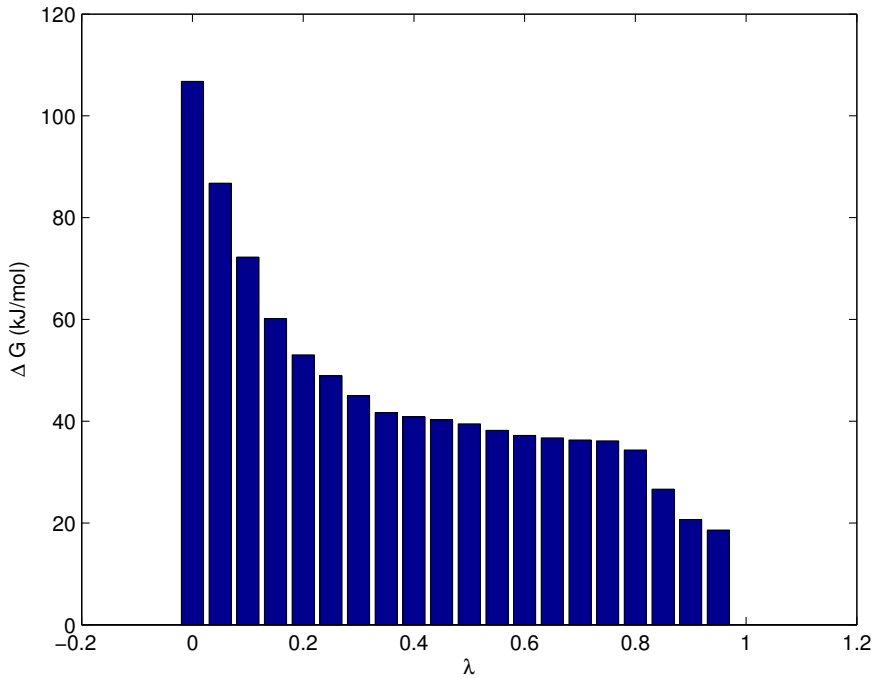


Figure 21: Free energy difference between neighboring values of λ for the bound barnase-barstar complex.

the barstar alone, the systems appeared to be equilibrated by the time the simulation reaches the production MD. In order to compare with the implicit water results, we had set up the thermodynamic cycle with the assumption of rigid binding. This meant that the barstar, when by itself, was not in an ideal conformation, but because it was restrained to not move about, the extra potential energy from being in a non ideal conformation should cancel out.

The barstar system's density was 994.0 $\frac{\text{kg}}{\text{m}^3}$, compared to the 1010 $\frac{\text{kg}}{\text{m}^3}$of the barnase-barstar system. Part of the reason that it decreased is because protein is denser than water. However, it is odd that the density is below that of water; even if waters were not incorporated correctly in the starting structure, the minimization should have caused the system to reach the correct density. The negative pressures observed at some of the λ values could also be an indication the system is held at too low a density[56] and wants to contract. Notice that if the density of the system was too low in the starting structure, the only opportunity to decrease in density would be during the energy minimization step.

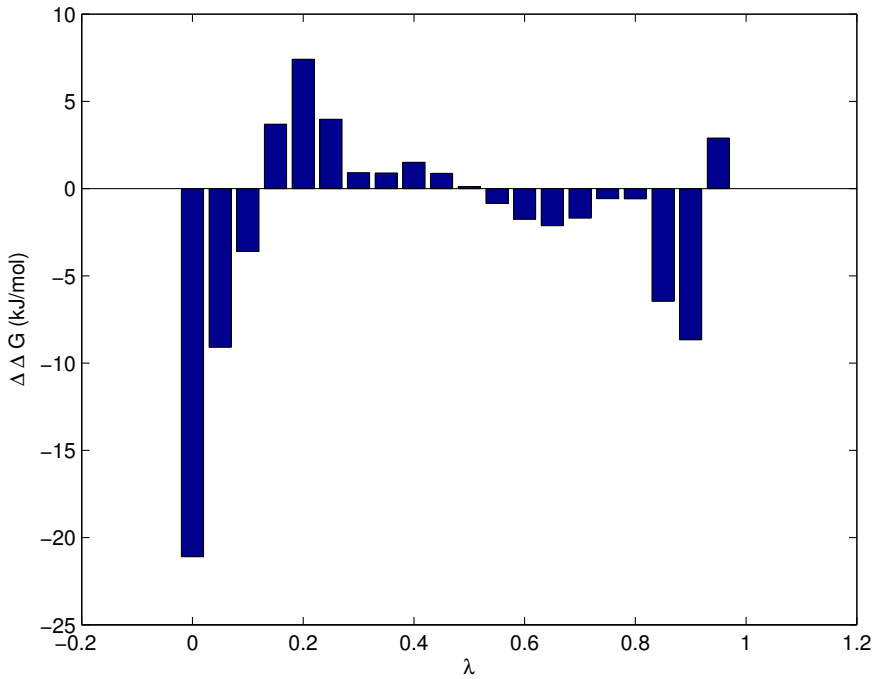


Figure 22: $\Delta\Delta G$ values between neighboring values of λ, where each $\Delta\Delta G$ is the difference between the free energy in the barnase-barstar complex and the free energy in the barstar alone at that λ. $\Delta\Delta G = \Delta G_{barnase-barstar} - \Delta G_{barstar}$.

The trend in the free energy differences between neighboring values of λ is expected to be quadratic, due to linear response theory. When a charge is doubled, the reaction field generated by the water in response to the charge also doubles, resulting in a quadratic change in solvation energy. The interaction energy change caused by the change in charge will be linear. Because a quadratic plus a linear change results in a total quadratic change, we expect a change in charge to impact the energy quadratically. However, looking at figure 21, notice that the relationship between charge and free energy is not quadratic – at low λ (charge close to 1), it appears quadratic, but not at high λ (charge close to 0). This seems to indicate that the linear response model does not hold perfectly at lower charges, an odd result that warrants further inspection.

4 Future work

During the energy minimization, the potential energy does not look fully minimized; part of the cause of this may be a too low density in the system. Originally, we had done two energy minimizations, one steepest descents and one L-BFGS. However, in GROMACS, only steepest descents is set up to handle restraints properly. This means we were unable to use a minimization method other than steepest descents to help reach a lower energy. We could allow the steepest descents minimization to run for more steps and see if this fixes the issue with the density.

Additionally, there could be problems with the starting structure, if it was not fully minimized before beginning the FEP. In addition to running a longer minimization, we could also let it equilibrate. We had attempted to ensure it was in an ideal configuration by running an MD simulation prior to the FEP, but as discussed earlier, this resulted in an altered interface. A potential cause of this could be that the force field used is not properly parameterized to keep the interface stable; I would like to test different force fields to see if there is one that would keep the interface stable and allow the starting structure to be in a

better configuration.

There are also some inconsistencies that need to be addressed. Prior to performing the FEP on the barstar system, we did not run an additional minimization to move the waters to energetically better places. As a result, the minimizations for the barstar system did not converge as well as for the barnase-barstar system. There were also some differences between the conditions used to equilibrate the barnase-barstar structure (to produce the structure with altered interface) and the conditions used during the course of the FEP; the barnase-barstar equilibration was done at 310 K, but the FEP was done at 300 K.

Although the phase overlap is not high in any of the FEP runs, the error due to this overlap is relatively low, on the order of a few kJ/mol. Running more simulations at intermediate λ values, particularly at low λ values where the phase overlap is worse, would be able to reduce this error even more.

Ideally, we would like to incorporate crowders into this system, and compare the $\Delta\Delta\Delta G$s to those in implicit water. Because the crowders we want to use are larger than normal atoms, there are no physical Lennard-Jones potential values we can use to motivate them. The crowders will be restrained in the same way the proteins are, in order to be comparable to the implicit water results. In our previous study,[30] we noticed that the placement of crowders has a large effect on the binding free energy. In order to achieve statistically significant results, it was necessary to average over many randomly generated crowder placements. To compare the explicit solvent results to implicit solvent results, we will have to compare structures containing crowders in the same locations. Because the structure used in this study is slightly different than the one used in the previous study, it will be necessary to repeat the implicit solvent experiments on this new structure. Then, FEP could be run on the same crowder configurations.

There are also several interesting parameters that we studied in implicit solvent and that would be interesting to study in explicit solvent. For example, higher salt concentrations were shown in implicit solvent to not change trends, but, on average, mute the effect of

crowders. I hypothesize that a similar effect will be found in explicit solvent; more salt will again mute the effect of the crowders. It would also be interesting to perform component analysis on the other four residues studied in our previous work.

The barnase-barstar complex was chosen because it binds intracellularly and therefore under crowded conditions. A search of the protein databank did not yield many complexes that unambiguously bound intracellularly and were well-suited for computational study. If other appropriate intracellular protein complexes could be identified, it would be interesting to perform these experiments on those complexes.

Eventually, it would also be interesting to understand how other elements of "reality" will affect binding free energy. For example, to bring in some elements of soft interactions, charged crowders could be used. A more computationally demanding experiment could also use actual protein shapes for crowders, instead of spheres.

In conclusion, this thesis lays the groundwork for studying the effect of crowding on the contribution of a particular residue's charges to the binding free energy in explicit solvent. We have done preliminary studies on the change in binding free energy associated with zeroing the charge on barstar's ASP35. Ultimately, we hope that this work will lead to a better understanding of the effects of crowding, which will in turn lead to better models of protein binding.

5 Bibliography

[1] Ellis, R. J.; Minton, A. P. *Nature* **2003**, *425*, 27–28.

[2] Chen, Z.; Zheng, K.-W.; Hao, Y.-h.; Tan, Z. *J Am Chem Soc* **2009**, *131*, 10430–10438.

[3] Cheung, M. S.; Klimov, D.; Thirumalai, D. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 4753–4758.

[4] Wang, Y.; Li, C.; Pielak, G. J. *J Am Chem Soc* **2010**, *132*, 9392–9397.

[5] Chen, E.; Christiansen, A.; Wang, Q.; Cheung, M.; Kliger, D.; Wittung-Stafshede, P. *Biochemistry* **2012**, *51*, 9836–9845.

[6] Homouz, D.; Stagg, L.; Wittung-Stafshede, P.; Cheung, M. S. *Biophysical Journal* **2009**, *96*, 671–680.

[7] Zhou, H.-X.; Rivas, G.; Minton, A. P. *Annu Rev Biophys* **2008**, *37*, 375–397.

[8] Batra, J.; Xu, K.; Qin, S.; Zhou, H.-X. *Biophys J* **2009**, *97*, 906–911.

[9] Christiansen, A.; Wang, Q.; Samiotakis, A.; Cheung, M. S.; Wittung-Stafshede, P. *Biochemistry* **2010**, *49*, 6519–6530.

[10] Lim, W. K.; Rosgen, J.; Englander, S. W. *Proc Natl Acad Sci U S A* **2009**, *106*, 2595–2600.

[11] Doan-Nguyen, V.; Loria, J. P. *Protein Science* **2007**, *16*, 20–29.

[12] Charlton, L. M.; Barnes, C. O.; Li, C.; Orans, J.; Young, G. B.; Pielak, G. J. *J Am Chem Soc* **2008**, *130*, 6826–6830.

[13] Miklos, A. C.; Li, C.; Sharaf, N. G.; Pielak, G. J. *Biochemistry* **2010**, *49*, 6984–6991.

[14] Harada, R.; Sugita, Y.; Feig, M. *Journal of the American Chemical Society* **2012**, *134*, 4842–4849.

[15] Predeus, A. V.; Gul, S.; Gopal, S. M.; Feig, M. *J Phys Chem B* **2012**, *116*, 8610–8620.

[16] Tanizaki, S.; Clifford, J.; Connelly, B. D.; Feig, M. *Biophys J* **2008**, *94*, 747–759.

[17] Radhakrishnan, M. L. *Theoretical Chemistry Accounts* **2012**, *131*, 1–24.

[18] Wang, Y.-T.; Wright, J. D.; Doudeva, L. G.; Jhang, H.-C.; Lim, C.; Yuan, H. S. *J Am Chem Soc* **2009**, *131*, 17345–17353.

[19] Lippow, S. M.; Wittrup, K. D.; Tidor, B. *Nat Biotechnol* **2007**, *25*, 1171–1176.

[20] Halle, B. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **2004**, *359*, 1207–1224.

[21] Tehei, M.; Franzetti, B.; Wood, K.; Gabel, F.; Fabiani, E.; Jasnin, M.; Zamponi, M.; Oesterhelt, D.; Zaccai, G.; Ginzburg, M.; Ginzburg, B.-Z. *Proceedings of the National Academy of Sciences* **2007**, *104*, 766–771.

[22] Williams, S. P.; Haggie, P. M.; Brindle, K. M. *Biophys J* **1997**, *72*, 490–498.

[23] Mastro, A. M.; Babich, M. A.; Taylor, W. D.; Keith, A. D. *Proceedings of the National Academy of Sciences* **1984**, *81*, 3414–3418.

[24] Foster, K. R.; Resing, H. A.; Garroway, A. N. *Science* **1976**, *194*, 324–326.

[25] Verma, P. K.; Rakshit, S.; Mitra, R. K.; Pal, S. K. *Biochimie* **2011**, *93*, 1424–1433.

[26] Hassan, S. A.; Steinbach, P. J. *The Journal of Physical Chemistry B* **2011**, *115*, 14668–14682.

[27] McGuffee, S. R.; Elcock, A. H. *PLoS Comput Biol* **2010**, *6*, e1000694.

[28] Sasmal, D. K.; Ghosh, S.; Das, A. K.; Bhattacharyya, K. *Langmuir* **2013**, *29*, 2289–2298.

[29] Despa, F.; Fernandez, A.; Berry, R. S. *Phys Rev Lett* **2004**, *93*, 228104.

[30] Qi, H. W.; Nakka, P.; Chen, C.; Radhakrishnan, M. L. *PLOS ONE* **in press**,

[31] Lee, L. P.; Tidor, B. *Nat Struct Biol* **2001**, *8*, 73–76.

[32] Lee, L. P.; Tidor, B. *Protein Sci* **2001**, *10*, 362–377.

[33] Green, D. F.; Tidor, B. *Proteins* **2005**, *60*, 644–657.

[34] Minkara, M. S.; Davis, P. H.; Radhakrishnan, M. L. *Proteins* **2011**, *80*, 573–590.

[35] Hendsch, Z. S.; Tidor, B. *Protein Sci* **1999**, *8*, 1381–1392.

[36] Carrascal, N.; Green, D. F. *J Phys Chem B* **2010**, *114*, 5096–5116.

[37] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *Journal of Computational Chemistry* **1997**, *18*, 1463–1472.

[38] Hockney, R.; Goel, S.; Eastwood, J. *Journal of Computational Physics* **1974**, *14*, 148 – 158.

[39] Buckle, A. M.; Schreiber, G.; Fersht, A. R. *Biochemistry* **1994**, *33*, 8878–8889.

[40] Darden, T.; Perera, L.; Li, L.; Pedersen, L. *Structure* **1999**, *7*, R55–60.

[41] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

[42] MacKerell, A. D. et al. *The Journal of Physical Chemistry B* **1998**, *102*, 3586–3616.

[43] Zwanzig, R. W. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.

[44] Bennett, C. H. *Journal of Computational Physics* **1976**, *22*, 245 – 268.

[45] Brünger, A. T.; Karplus, M. *Proteins: Structure, Function, and Bioinformatics* **1988**, *4*, 148–156.

[46] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187–217.

[47] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

[48] Šali, A.; Blundell, T. L. *Journal of Molecular Biology* **1993**, *234*, 779–815.

[49] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–447.

[50] Berendsen, H.; van der Spoel, D.; van Drunen, R. *Computer Physics Communications* **1995**, *91*, 43 – 56.

[51] Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *The Journal of Physical Chemistry A* **1999**, *103*, 3596–3607.

[52] `http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/index.html`.

[53] Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

[54] The Mathworks, Inc., Natick, MA.

[55] Sheinerman, F. B.; Honig, B. *J Mol Biol* **2002**, *318*, 161–177.

[56] Rapaport, D. C. *The Art of Molecular Dynamics Simulation*; Cambridge University Press, 2004.

6 Appendix

This appendix contains the contents of the implicit water paper,[30] which has been accepted for publication in *PLOS ONE*.

The Effect of Macromolecular Crowding on the Electrostatic Component of Barnase—Barstar Binding: A Computational, Implicit Solvent-Based Study

Authors and Affiliations:

Helena W. Qi[1], Priyanka Nakka[1], Connie Chen[1], and Mala L Radhakrishnan[1]

[1]Department of Chemistry, Wellesley College, Wellesley, Massachusetts, United States of America

Corresponding Author:

Mala L. Radhakrishnan,

Department of Chemistry, Wellesley College

106 Central Street

Wellesley, MA 02481

(781)283-2981 (phone), (781)283-3642 (fax), mradhakr@wellesley.edu

Abstract

Macromolecular crowding within the cell can impact both protein folding and binding. Earlier models of cellular crowding focused on the excluded volume, entropic effect of crowding agents, which generally favors compact protein states. Recently, other effects of crowding have been explored, including enthalpically-related crowder—protein interactions and changes in solvation properties. In this work, we explore the effects of macromolecular crowding on the electrostatic desolvation and solvent-screened interaction components of protein—protein binding. Our simple model enables us to focus

exclusively on the electrostatic effects of water depletion on protein binding due to crowding, providing us with the ability to systematically analyze and quantify these potentially intuitive effects. We use the barnase—barstar complex as a model system and randomly placed, uncharged spheres within implicit solvent to model crowding in an aqueous environment. On average, we find that the desolvation free energy penalties incurred by partners upon binding are lowered in a crowded environment and solvent-screened interactions are amplified. At a constant crowder density (fraction of total available volume occupied by crowders), this effect generally increases as the radius of model crowders decreases, but the strength and nature of this trend can depend on the water probe radius used to generate the molecular surface in the continuum model. In general, there is huge variation in desolvation penalties as a function of the random crowder positions. Results with explicit model crowders can be qualitatively similar to those using a lowered "effective" solvent dielectric to account for crowding, although the "best" effective dielectric constant will likely depend on multiple system properties. Taken together, this work systematically demonstrates, quantifies, and analyzes qualitative intuition-based insights into the effects of water depletion due to crowding on the electrostatic component of protein binding, and it provides an initial framework for future analyses.

Introduction

It is believed that up to 40% of the cellular volume is occupied by macromolecules[1], making the cell a crowded place. Nevertheless, many *in vitro* experiments and computational studies model protein processes in a vast "sea" of aqueous

solvent. To build better models of such processes, it is crucial to better understand the effect of cellular crowding on the physical determinants of protein folding and binding. While more attention has been given to these effects in recent years, reviews of crowding effects span multiple decades[2-9]. Experimental work has shown that crowding can cause a thermodynamic favoring of compact states – folded, bound, or aggregated states of proteins[10-13] – and could favor compaction of unfolded states as well[14,15], although sometimes certain effects were found to be small or even reversed[16,17], likely because of enthalpic interactions between crowding agents and the proteins being studied[18]. Nevertheless, even small, subtle effects could have important implications for aggregation associated with neurodegenerative diseases[10,19]. Crowding has also been experimentally shown to change the preferred conformations of protein and DNA systems[20-25] and to alter drug—target interactions or affinities[26-28]. Finally, macromolecular crowding may slightly[16,29] or more greatly affect association rate kinetics[30] and reaction mechanisms[31,32].

Theoretical and computational studies have provided great insight into the physical bases for observed effects due to macromolecular crowding. Many thermodynamic studies to date have focused on the entropic "excluded volume" effect, in which crowding lowers the available cellular volume, thus lowering the entropy of noncompact states more than that of compact states, leading to a relative free energy stabilization of compact states. This effect was shown to have measurable consequences in theoretical and computational studies[33-36]. More recently, it was shown that favorable interactions between less compact states and the crowders could cancel out this effect or dominate over it[37-39], demonstrating not only that the physical properties of the crowders are important, but also

that crowding could significantly affect the enthalpic component of the binding free energy in addition to the entropic component. The subtle interplay between multiple energetic components as well as dynamical effects have been considered via molecular dynamics simulations of proteins within a crowded environment[37,38,40,41]. These and other time-dependent simulations[42,43] have also provided insight into the association rates of proteins within the cellular milieu.

There have been relatively few studies that focus on how crowding affects the *electrostatic* component of protein—protein interactions and their solvation energetics. As a reasonable hypothesis, crowding can both affect the hydration dynamics of water[44] and deplete the number of polarizable water molecules surrounding the proteins, thereby potentially descreening their electrostatic interactions relative to the infinite dilution limit (i.e., the uncrowded case). While crowding has been incorporated into electrostatic models via a screened Coulomb potential-based implicit solvent model[45] and a lowered effective solvent dielectric constant[46], to our knowledge, only very recent work has probed more specifically to study how crowding affects electrostatic interactions within a solvated medium[47,48]. Such work demonstrated that it may be possible to capture certain electrostatic effects of crowding by a lowered solvent dielectric constant, a result that supports other work suggesting that the observed dielectric constants within cellular environments may be quite lower than that of water[49-53]. Specifically, Harada *et al*.[47] found via explicit solvent molecular dynamics simulations that water mobility was hindered in a crowded environment, providing one physical mechanism for this lowered dielectric constant. However, as they note, another mechanism for a lowered dielectric constant may stem from the fact that crowding depletes bulk water from around molecules,

an idea that was explored further in an implicit model study[48]. It is this latter mechanism that provides the focus of the current study, although here, we extend this idea to study protein—protein binding.

This work uses simplified models to study how water depletion due to crowders can alter electrostatic binding free energies between proteins. We use the barnase—barstar protein complex as a model system, as it has been shown previously[54,55] that electrostatic interactions play a crucial role in their interaction, and it has also been used in previous studies investigating crowding or similar phenomena[35,45]. While a more realistic model may use explicit solvent and actual proteins as crowding agents, we wished to separate out electrostatic effects due to water depletion from other electrostatic effects, such as loss of mobility of individual water molecules or electrostatic interactions with crowder molecules. To that end, our study uses spherical, uncharged model crowders within an implicit solvent, and electrostatic free energies are computed through obtaining potentials via the Poisson Equation (or the Linearized Poisson-Boltzmann equation, if applicable). To again focus on the water depletion effect in a controlled manner, we assume rigid binding, although we recognize that crowding may affect protein conformations[48]. Our thermodynamic cycle allows us to separately quantify the effects of crowding on desolvation and on solvent-screened interaction. The use of simple model crowders enables us to systematically study these effects as a function of crowder density and size. Adequately sampling crowder locations to get proper Boltzmann-weighted distributions of states would be computationally infeasible, and so we limited our results to simple averages over 50 randomly-generated crowder placements in the bound and unbound states per data point, especially since Boltzmann-weighting based only on

electrostatic solvation energies may be less realistic than assuming that other factors can also contribute to crowder placement.

We find that on average, crowding lowers desolvation penalties and amplifies solvent-screened interactions, stabilizing favorable interactions and destabilizing unfavorable ones. This effect is more pronounced when crowder size is reduced, assuming a standard-size water probe radius within the continuum model. The mean stabilization or destabilization of solvent-screened interactions was robust to the specific placement of the random crowders, but the average desolvation effects were not, with very large standard error values. While an overall reduced dielectric constant may capture average water depletion effects, there may be system specific conditions that lead to uncertainty in the mean effect of crowder placement as a simple function of crowder density and size. Finally, we show that crowding can differentially affect the electrostatic contributions of individual protein residue side chains toward binding, with the relative effects on desolvation and interaction depending on the residue's environment. This suggests that crowding could affect the consequences of specific mutations on binding, as well as the role that certain residues or binding "hot spots" play in varied cellular environments. While these results may qualitatively agree with intuition, our goal is to provide a systematic, controlled demonstration and quantitative analysis of these effects. Moreover, the methods used here provide experimentally testable hypotheses and an initial framework for understanding the role of crowding in modulating electrostatic interactions in protein—protein binding that can be built upon in future work.

Materials and Methods

Structure Preparation

Studies used a 2.0 Å resolution crystal structure of barnase complexed with a Cys -> Ala (40,82) double mutant of barstar (PDB ID 1BRS)[56]. The asymmetric unit consisted of 3 model complexes; the complex corresponding to chains A and D were used in this study. Crystallographic water molecules greater than 3.3 Å from either binding partner or with fewer than three potential hydrogen-bonding interactions with protein were removed. The remaining 17 water molecules were assigned to either protein partner based on proximity and hydrogen-bonding contacts. The amide groups of asparagine and glutamine and the imidazole group of histidine were flipped as necessary and the tautomerization states of histidine were assigned based on manual inspection of possible hydrogen bonding with surrounding residues. The two N-terminal residues of barnase and residues 64 and 65 of barstar were not resolved in the crystallographic experiment, and neighboring residues were patched with acetyl or N-methylamide groups. Hydrogens were modeled onto the structure with the HBUILD[57] functionality in CHARMM[58], using the CHARMM22 force field[59] and the TIP3P water model[60]. Patches and missing side chain density were added via CHARMM and were energy minimized.

Crowder Placement

Bound and unbound states in each binding free energy calculation were crowded separately. A box was created to contain both the protein complex (or each unbound state) and the model crowders, such that the box "walls" were each 70 Å from the most extreme (i.e., maximal and minimal) x, y, and z protein coordinates. The dimensions of the box were approximately 190x190x190 Å. Spherical crowders of either specified or random radii (up

to 25 Å, roughly the size of the barnase—barstar complex) were added sequentially, and each potentially new crowder was accepted if it did not (1) overlap in space with any existing crowder or protein molecule, (2) partially or totally fall outside the total box volume, or (3) cause the volume density of crowders to be higher than the desired value. The volume density of crowders was calculated as the ratio of the total volume of the crowders to the originally available volume (i.e., volume not taken up by the protein(s)). Fig. 1 shows sample, random crowder placements around the bound state at denoted specifications. Preliminary analyses showed that one consequence of our crowder placement method is a depletion of crowder density at the system's extreme edges; future efforts to place crowders could adopt a strategy leading to more even placement throughout the entire system volume.

Continuum Electrostatics Calculations

A single-grid red-black successive over-relaxation finite-difference solver (M.D. Altman and B. Tidor, unpublished)[61] of the Poisson/Linearized Poisson Boltzmann Equation, distributed with the Integrated Continuum Electrostatics (ICE) software package (D.F. Green, E. Kangas, Z.S. Hendsch, and B. Tidor, Massachusetts Institute of Technology Technology Licensing Office), was used to solve for the electrostatic potentials of both crowded and uncrowded systems. Unless otherwise noted, a probe radius of 1.4 Å was used to define the molecular surface for the dielectric boundaries. Likewise, unless otherwise noted, a dielectric constant of 4 was used for all spherical crowders and protein atoms, and the solvent was modeled using a dielectric constant of 80. Potentials were solved on a 491x491x491 grid. A three-tiered focusing procedure was used, in which the system (the complex and all crowders) occupied 23%, 92%, and 184% of the grid. At the

lowest focusing, the regions beyond the entire system were modeled as dielectric 80 and screened Coulombic (or Debye-Huckel, in cases of non-zero ionic strength) boundary conditions were used. Zero-radius dummy atoms were placed at identical extreme points of every run to maintain equal grid resolution for all states. At the highest focusing, this grid spacing yielded a resolution of approximately 4.6 grids/Å, and the grid was centered on barstar within the large system (for a small subset of runs, the grid was centered on a particular atom within the interfacial barstar Asp39 residue). PARSE radii and charges[62] were used. The ionic strength was set to zero except when implicit salt was modeled at a concentration of 0.145M and a Stern layer of 2 Å was used. Due to memory limitations, runs with nonzero ionic strength were solved on a 401 x 401 x 401 grid, and to assess the effect of ionic strength, were compared only to other runs at the same grid resolution.

Potentials were solved for both the bound and unbound dielectric boundaries upon charging up one binding partner at a time. By multiplying (one-half) the potential differences due to charges on a given partner by the charges on that partner, desolvation penalties were obtained, and by multiplying the potentials due to charges on one partner by the charges on the other partner, solvent screened interactions were obtained[63] (Fig. 2).

Model charge variation

The monopole on each binding partner was changed by adding or subtracting random charge values of maximum magnitude 0.1e to randomly selected atoms within the partner until the desired overall monopole was reached. No single atom was allowed to have an overall charge magnitude greater than 0.85e. To test the robustness of the results, monopoles were changed by starting both with the original charge distribution and from a

structure in which all the charges were set to zero. Here we show only the results produced by starting with the original barnase-barstar charge distribution.

Component Analyses

To quantify the contributions of selected residues toward the electrostatic component of binding in the presence and absence of model crowders, the partial atomic charges on the side chain of a given residue were all set to zero and the binding free energy re-evaluated, in a similar manner to component analyses in previous work on both protein and small molecule systems[55,64-68]. The effect of zeroing out the side chain was then computed via:

$$\Delta\Delta G_{res} = \Delta G_{zeroed} - \Delta G_{orig}$$

A positive value of $\Delta\Delta G_{res}$ implies that a residue's side chain contributes favorably toward the electrostatic component of binding, as zeroing out its charges worsens binding. The desolvation and interaction components of $\Delta\Delta G_{res}$ were computed by directly subtracting the desolvation and interaction components of the binding free energies between the system with zeroed charges and the original system, respectively.

Component Analyses of residue groups within barstar

For analyses in which charges of groups of residues were zeroed, groups were determined by calculating the solvent accessible surface area (SASA) of residues within each partner (assuming associated water molecules are considered residues and not bulk solvent) in the bound and unbound states. CHARMM was used to calculate SASA, using a 1.4 Å -radius probe and the CHARMM22 force field. Residues with non-zero burial upon binding were classified as either highly buried or peripheral depending on whether more or less than

50% of their unbound SASA remained in the bound state. Non-core residues were classified as either surface exposed or partially exposed depending on whether they have more or less than 50 Å^2 SASA in the unbound state. Here, the charges of both side chain and backbone atoms were set to zero so that the union of all atoms considered was the entire barstar protein (and associated explicit water molecules).

Data Analysis and Visualization

Figures of protein molecules and model crowder systems were generated using VMD[69]. All plots and data analyses were performed using Matlab (The Mathworks, Inc. Natick, MA).

Results

To assess the effect of water depletion due to crowding on the electrostatic component of protein—protein binding, binding free energies were computed in the presence and absence of model crowders. To model the crowded states in a controlled fashion and focus on water depletion, spherical, uncharged "crowders" were randomly placed around the bound and unbound state proteins at specified densities (Fig. 1). The effect of crowding on the electrostatic component of the binding free energy was quantified as the difference between the electrostatic binding free energies in the presence and absence of crowders:

$$\Delta\Delta G_{crowding} = \Delta G_{bind,crowded} - \Delta G_{bind,uncrowded}$$

A negative $\Delta\Delta G_{crowding}$ means that crowding lowers the electrostatic binding free energy (i.e., favors binding, all other components equal). With our model, $\Delta G_{bind,elec,uncrowded}$ was found to be 0.5 kcal/mol, suggesting that the electrostatic component of binding in this system (in

pure aqueous solvent) is neither strongly favorable nor unfavorable, in qualitative agreement with previous work using quantitatively different parameters[70]. Given that the electrostatic binding free energies between proteins are generally quite unfavorable with models using an internal dielectric constant of 4[71], our value supports the accepted view that electrostatics play an important role in this system.

Binding free energy contributions were broken into desolvation and interaction components (Fig. 2). The free energy cost upon binding to remove solvent interactions with barstar (considered the "ligand") is denoted the ligand desolvation penalty (LDP), and was found to be 41.7 kcal/mol for the uncrowded system. The energetic cost upon binding to remove solvent around barnase (the "receptor") is termed the receptor desolvation penalty (RDP, 37.2 kcal/mol when uncrowded). Finally, the solvent-screened interaction between the partners (int) was also quantified (-78.4 kcal/mol when uncrowded).

On average, crowding lowers desolvation penalties and amplifies interactions

Figure 3 is a graph of $\Delta\Delta G_{crowding}$ as a function of crowder radius (bars grouped by bottom axis) and crowder volume density (top axis). In the rightmost set of bars, crowder radii vary within each system from 5-25 Å (the largest spheres were therefore approximately the size of the protein complex). Total $\Delta\Delta G_{crowding}$ values are broken up into contributions due to changes in barstar's desolvation penalty (LDP, blue), barnase's desolvation penalty (RDP, green), and solvent-screened interaction (int, red). Each bar is the result of 50 random trials, with average values +/- standard error (not standard deviation) shown for each contribution.

Figure 3 shows that on average, $\Delta\Delta G_{crowding}$ was negative for all crowder densities and radii, although generally, the effects were more pronounced at higher crowder densities and smaller crowder radii. Moreover, the changes in all contributions (LDP, RDP, and int) were generally negative on average, in this system. This result makes intuitive sense – in a crowded environment, each unbound state is already partially desolvated by crowders, with some crowders potentially occupying the same space in the unbound state as the binding partner does in the bound state. Hence, there may be less solvent displaced near the binding interface upon binding in the crowded system when compared to an uncrowded one, resulting in a reduced desolvation penalty on average. Moreover, the bound state is also partially desolvated due to the crowding, resulting in less solvent screening and more amplified interactions between the two partners. Because the interactions in this complex are favorable in general, amplifying them would increase their favorability.

The average effects seen in Fig. 3 are qualitatively similar to what one might obtain using a lower solvent dielectric constant. Previous work has modeled aspects of crowding via the use of a lower "effective" solvent dielectric constant[37,38,46,48], and experimental evidence suggests that a dielectric constant can be characterized for the cytoplasm[51,53] through measuring shifts in emission wavelength maxima of fluorescent probes due to the polarity of the microenvironment. This observed constant likely is a macroscopic average accounting for both the loss of water mobility and water depletion (and potentially other effects), the first of which is not accounted for in the present study. Nevertheless, it is instructive to measure the effects of a lowered, effective solvent dielectric on protein—protein binding. Figure S1 shows $\Delta\Delta G$ values (relative to a solvent dielectric constant of 80)

for the desolvation and interaction components of barnase-barstar binding as a function of solvent dielectric constant. In addition, Table 1 shows numerical data using two potential values of solvent dielectric constant – an experimentally obtained value of 21.9[53] and the value of 55, similar to values found from explicit simulations at 30% crowder volume density, to model solely the effects of hindered water mobility[47]. A dielectric constant of 21.9 produced ΔΔG values that were several times more pronounced (Table 1) than the results obtained using explicit crowders (Fig. 3), but this may be because the experimentally-obtained constant would account for not only water depletion, but also hindered water mobility and other possible effects of crowding. A dielectric constant of 55 again produced more pronounced results than using explicit crowders within a dielectric 80 medium, although the effects were more quantitatively similar to our explicit crowding simulations (~1kcal/mol difference in ΔΔG for desolvation components and ~5 kcal/mol difference in ΔΔG for interaction, at a 30% crowding density and varied radius, Table 1). Again, differences could be due to the fact that this value was found to account for hindered water mobility and not water depletion.

The qualitative trends seen with lowered dielectric constants (Fig. S1) were similar to the trends found in this work for either increasing crowder volume density or decreasing radius, although for a given crowder radius and volume density, there may not exist an effective dielectric constant that provides quantitative agreement. Perhaps a "long-range" dielectric constant cannot model the full effect of hydration immediately surrounding each macromolecule; in a heterogeneous environment, the dampening of the electric fields due to a small amount of highly polar water might not be captured by an average, low macroscopic dielectric constant and therefore, effects of crowding may be

overestimated. Nevertheless, one potential solution, similar to what was done in work by Harada *et al.*[38], is to use a slightly lower dielectric constant to account for the loss of water mobility and explicitly model crowders to account for water depletion. Future work could also involve effective medium theory approaches to estimate effective dielectric constants of this composite environment as a function of crowder size and shape[72].

The relatively small standard error for interaction indicates that the mean stabilization due to the further descreening of interactions relative to infinite dilution is fairly robust to the ensemble of states sampled; there is little uncertainty in the mean effect. However, the large standard error for both desolvation contributions in all ensembles indicates great uncertainty in the mean reduction of desolvation penalties due to random crowder placement. As desolvation penalties depend strongly on the level of direct solvent exposure of charged or polar interfacial groups, it makes sense that they will be very sensitive to precise crowder placement. Interaction energies, on the other hand, are more long-ranged, except for interfacial interactions (and these are fairly unaffected by crowders in the bound state anyhow), and are therefore far less sensitive. The large standard error due to desolvation, by definition, implies an even larger standard deviation and therefore a huge amount of variability between trials, which suggests the necessity of thorough sampling. Currently, it is computationally infeasible to thoroughly sample all relevant crowder configurations. Preliminary attempts to use Boltzmann-weighting to more heavily account for lower-energy states by obtaining partition functions from each set of 50 sampled configurations resulted in similar qualitative trends to those shown in Fig. 3 (data not shown).

Our results suggest that the effects of crowding on water depletion are most pronounced at a given crowder volume density when the crowders are small, although large standard errors confound the robustness of this result, especially for desolvation. Presumably, very small molecules can more closely approach the irregular surface of a protein, more substantially desolvating it in its unbound state and more effectively descreening its interactions with a partner in the bound state relative to infinite dilution. Analyses of our model crowded systems showed that the minimum distance of approach between any one crowder and the proteins increases on average as the crowder radius increases (Figure S2), in support of this hypothesis.

It is plausible that aspects of this observed trend could be dependent on the use of a standard, nonzero-sized (here, 1.4 Å) "probe" used to generate the molecular surface in continuum models. The water-sized probe is intended (as standard practice) to approximately account for the nonzero size of discrete water molecules and the inability of "actual" water molecules to occupy cavities and crevices smaller than their size. A consequence of this model feature is that low-dielectric regions will be larger than the actual volume occupied by model crowders and protein, and this difference will likely be greater for systems with smaller-radius crowders due to the likelihood that they often closely approach each other and the protein.

To test this hypothesis, we redid a subset of the calculations shown in Figure 3 using a probe radius of zero to generate the molecular surface. The results are shown in Figure S3. Desolvation penalties were still reduced on average and interactions amplified, but as expected, the quantitative effects were now often ~50-75% less pronounced ($\Delta\Delta G_{crowding}$ = ~2 kcal/mol or less). Additionally, the dependence of the desolvation effects on radius was

not apparent (although they did not appear to be statistically significant even with a standard probe radius). However, the average effect on the interaction component still strengthened overall as the crowder radius decreased, suggesting some robustness to the observation that smaller crowders may have greater impact. While it is standard practice to use a probe radius of 1.4 Å[73,74], results using a continuum model can be sensitive to this feature[74,75]. Our results demonstrate this limitation, specifically when modeling crowding effects using a continuum approach.

Even with the "standard" probe radius of 1.4 Å, at radii that more accurately model small proteins (20-25 Å), the mean effects on electrostatic interaction were found to be modest, but still significant on average, especially at higher crowding densities. These data suggest that the effects of crowding on electrostatics could be sensitive to the precise distribution of molecular sizes within the cell, and that it might be not be crowding due to proteins but rather, due to smaller metabolites and peptides that most greatly affects the electrostatic component of binding. We note that the trends for radii are curtailed here due to missing data at higher crowder densities and larger radii. Because of our purely random, sequential crowder placement, it became geometrically impossible to satisfy all constraints noted in the Methods when both crowder size and desired volume density were large. Future work can attempt to explore this region of property space while still maintaining a purely random crowder placement within the noted constraints.

Taken together, these results show that on average, the effects of crowding on electrostatic interactions can vary as a function of both crowder volume density and size, but desolvation effects are highly sensitive to crowder placement. To qualitatively account for crowding effects due to water depletion, therefore, it may be expedient to use an

effective lowered solvent dielectric constant. Our work supports the idea that such a constant is likely to be specific to crowding volume fraction[47] and the distribution of crowder radii, and additional parameters may be needed to capture system-specific variations due to various arrangements of crowders.

In addition to the varied probe radius size discussed above, a subset of data was obtained under other different model conditions, to gauge the robustness of our results to parameters and physical conditions. First, we varied the internal dielectric constant used for both protein and model crowders. For maximal control, the precise locations of crowders in the bound and unbound states of the 50 trials were maintained in calculations with different dielectric constants in one set of runs, and allowed to vary in another set. Results here used a varied crowder radius at a volume density of 30%. With an internal dielectric constant of 1, results were qualitatively similar to those with an internal dielectric constant of 4 when controlling for crowder placement and quantitatively more pronounced on average, especially for desolvation penalties (Table 1, $\varepsilon_{in} = 1$, same"). However, standard errors were much larger, which may explain the difference in $\Delta\Delta LDP_{crowding}$ between trials in which the same crowders were used and when random crowders were used (Table 1, "$\varepsilon_{in} = 1$, random").

To understand how the presence of electrolytes could modulate the effect of crowding, data were gathered including implicit mobile ions at a concentration of 0.145M through obtaining potentials via the linearized Poisson-Boltzmann equation. Again, we used a crowder volume density of 30% and randomly varied crowder radii, although all relevant runs with and without mobile ions were done at a somewhat lower grid resolution due to memory limitations when modeling salt (see Methods). We obtained qualitatively

similar results when the solvent contained implicit, mobile ions, although the average lowering of the LDP, RDP, and especially int, were not as pronounced (Table 1).

If crowders descreen interactions relative to infinite dilution, they should amplify both attractive and repulsive interactions. To show this, we computationally modified the charge distributions on both barstar and barnase to vary their monopoles (see Methods). Of course, such charge distributions are not realistic, but they allow for a controlled, systematic study on how a system's charge distribution may affect its molecular recognition profile in a crowded environment. Figure 4 shows the average change in LDP, RDP, and int for three modeled pairs of monopoles – in which the partners either had opposite, large-magnitude monopoles (+/-10e), no net monopole, or the same, large-magnitude monopole (+10e). Each bar is the average of 50 trials in which crowders of varied (5-25 Å) radius were used at a 30% volume density. The average effect of crowding on desolvation penalties was similarly stabilizing in all three cases, but the average effect on interactions is markedly different in the three cases. As expected, crowding greatly destabilized the (+10/+10) interaction and greatly stabilized the (+10/-10) one. This suggests that binding partners' overall monopoles can affect how they interact with partners in a crowded environment, although this effect is mediated more by interactions rather than the desolvation component.

Crowding can differentially affect electrostatic contributions of side chains toward binding

Many protein—protein interactions have been shown to be mediated by one or more polar or charged residues or "hot-spots"[76-79]; such residues can be elucidated by experimental mutagenesis studies (e.g., alanine scanning) or through computational analyses.

Presumably, if the overall electrostatic binding free energy can be modulated by the level of environmental crowding, as the model above suggests, then this implies that the specific contributions of individual residues toward that interaction can also be altered, but the nature of the alteration may depend on the properties of each residue.

To explicitly demonstrate, quantify, and better understand this intuitive idea, we began with the original (unaltered) charge distribution of the complex and quantified the electrostatic contribution of selected barstar residues toward the binding free energy by computationally setting the original partial atomic charges on a given side chain to zero and re-evaluating the binding free energy to obtain a $\Delta\Delta G_{res}$ (see Methods); this procedure was done both in the presence of crowders (the 50 trials used in the original analyses were used to obtain an average $\Delta\Delta G_{res}$) and in the absence of crowders. Consequently, we can define a $\Delta\Delta\Delta G_{res,crowding}$ that quantifies the effect of crowding on a residue's contribution toward the binding free energy:

$$\Delta\Delta\Delta G_{res,crowding} = \Delta\Delta G_{res,crowded} - \Delta\Delta G_{res,uncrowded}$$

A positive $\Delta\Delta\Delta G_{res,crowding}$ means that a residue contributes *more* favorably (or less unfavorably) toward binding in the presence of crowding than in its absence.

In this study, we chose to calculate $\Delta\Delta\Delta G_{res,crowding}$ for five barstar residues whose side chains were previously shown to contribute significantly toward the electrostatic component of binding free energy[55]: Tyr29, Asp35, Asp39, Thr42, and Glu76. Figure 5a is a graph of the $\Delta\Delta\Delta G_{res,crowding}$ for each of these residues, broken up into barstar desolvation (LDP) and interaction (int) components (there is no change in the desolvation of barnase, RDP, as only charges on barstar were changed to zero). On average, the charged side chains contributed even more favorably in the presence of crowding, although the

effect was quite small, with an average $\Delta\Delta\Delta G_{res,crowding}$ of only tenths of a kcal/mol. The contributions were not significantly changed on average for the two polar side chains studied.

Interestingly, the desolvation component of $\Delta\Delta\Delta G_{res,crowding}$ was altered more on average for Asp35 and Asp39, whereas the interaction component was altered more on average for Glu76. We hypothesize that the different mechanisms of altering $\Delta\Delta\Delta G_{res,crowding}$ is due to where these residues lie relative to the binding interface (Fig. 6). Both Asp35 and Asp39 are interfacial and highly buried upon binding, and so crowding may more greatly affect their desolvation penalties, by partially desolvating them already in the unbound state. Glu76, however, is more peripheral to the interface and so it remains more solvent exposed upon binding – this implies that crowding could more greatly impact the solvent-screening of its interactions in the bound state.

To further explore the idea that crowding might affect residue-based contributions differently, we grouped barstar residues based on both level of surface exposure and degree of burial upon binding (see Methods). Then, we zeroed out the charges simultaneously on all residues in each group (including both side chain and backbone) to determine $\Delta\Delta G_{res}$ for that group. This was done both in the presence and absence of crowding to obtain a $\Delta\Delta\Delta G_{res,crowding}$ (using the 50 trials used in the original analyses). Indeed, surface residues that are highly buried upon binding showed the largest desolvation component of $\Delta\Delta\Delta G_{res,crowding}$ values (Fig. 5b), while surface residues that are peripheral to the interface (i.e., only partially buried upon binding) showed the largest interaction component of $\Delta\Delta\Delta G_{res,\ crowding}$. Interestingly, $\Delta\Delta\Delta G_{res,crowding}$ of surface residues with no burial upon binding (i.e., distal from the interface) was negative; here, crowding

makes these residues contribute *more unfavorably* toward binding. This result may be due to the dominating effect of the monopoles of distal groups; the monopoles on our model of barnase (+1) and the collection of distal, surface exposed residues on barstar (+3) have the same sign. The same trends are found when one controls for the number of residues in each group by finding the average $\Delta\Delta\Delta G_{res,crowding}$ per residue in each group (Fig. S4). These results explicitly demonstrate that electrostatic contributions – and therefore perhaps mutational energies – can be predictably altered in an environmentally-dependent way for residues in a crowded environment.

Discussion

In this work, we used simplified models to investigate the effect of macromolecular crowding on the electrostatic component of protein—protein binding free energy via water depletion. We found that for proteins with favorable electrostatic interactions, crowding can enhance the relative favoring of the bound state due to lowered desolvation penalties and enhanced interactions. For proteins with potentially unfavorable interactions, there may be opposing effects. The effects of crowding on desolvation were highly sensitive to crowder placement – yielding far more uncertainty in the mean effect on desolvation than in the mean effect on the interaction component.

Our results can potentially provide experimentally-testable hypotheses. For example, one could experimentally study the effect of monopole-changing yet relatively isosteric (e.g., Asn→Asp) interfacial and peripheral mutations on protein—protein binding in crowded and uncrowded environments to see if crowding affects their relative contributions as predicted.; these experiments can be bolstered by varying ionic strength

to highlight the interaction component of binding over desolvation components. Experimental tests would likely combine the effects of crowding due to both water depletion and lowered solvent mobility, so experimental results should reflect the predictions in this work in combination with other computational predictions[47].

The importance of crowder size was studied in a previous computational study that focused on the excluded volume effect of crowding on the binding of the barnase—barstar complex[35]. Like our study, it was also found that smaller crowders had a larger effect, but for a different reason – at a given volume density, smaller crowders left smaller voids for the proteins to occupy, lowering the available volume. This effect was confirmed in another study, and it was also shown that the ratio between crowder size and protein size is important[12]. Thus, smaller crowders may have a bigger impact for multiple reasons – by their excluding more volume and by their ability to more closely approach proteins to desolvate them and descreen their electrostatic interactions relative to infinite dilution.

We also demonstrated that crowding can differentially affect the relative contributions of residues toward binding. That these changes can be dominated by different phenomena (desolvation vs. interaction) could provide avenues for rational, environmentally-dependent design tasks.

This study provides a useful framework on which to build in future studies. With adequate computational resources, larger-sized model crowders and overall crowded volumes could be explored. Elements of "reality" can be added individually, in turn, to understand the effect of each on the binding free energy. Such elements include using actual protein shapes for the crowders (crowder shape has been shown to affect changes in folding and binding free energies[12,80]) as well as protein charge distributions to include

direct enthalpic crowder interactions, which have been shown to be important for protein stability and conformation[18,48]; it would be interesting to quantify their precise effects on protein—protein binding. Another future goal is to increase the sampling of crowder configurations and potentially the conformational states of the binding partners, to allow for Boltzmann-weighted averages through Monte Carlo or dynamic simulations. In this study, the costs of Poisson-based models on such large systems prohibited exhaustive sampling (each binding free energy calculation took ~0.5 day of CPU time and > 1GB RAM with current resources).

To also account for the altered mobility of water molecules due to crowding, explicit solvent simulations are necessary, and have been previously attempted[38,47], although rigorously analyzing such effects on the energetics of specific protein—protein binding has yet to be done, to our knowledge. Given the potential computational cost of such studies, alchemical transitions[81,82] of individual residues (i.e., component analysis) or small molecule—protein binding systems may be good starting points.

In this study, we demonstrated and systematically explored the idea that macromolecular crowding can affect the electrostatic component of the free energy of binding between proteins through depleting regions of high dielectric water. Our results highlight yet another example of how environmental effects can have a quantitative and potentially qualitative impact on molecular recognition and should therefore be considered in both the analysis and the rational design of biomolecular systems.

Acknowledgements

The authors thank Michael Altman, Jaydeep Bardhan, David Green, and Bruce Tidor for helpful discussions and/or software.

References

1. Ellis RJ, Minton AP (2003) Cell biology - Join the crowd. Nature 425: 27-28.
2. Zimmerman SB, Minton AP (1993) Macromolecular crowding - Biochemical, biophysical, and physiological consequences. Annual Review of Biophysics and Biomolecular Structure 22: 27-65.
3. Ellis RJ (2001) Macromolecular crowding: Obvious but underappreciated. Trends in Biochemical Sciences 26: 597-604.
4. Hall D, Minton AP (2003) Macromolecular crowding: Qualitative and semiquantitative successes, quantitative challenges. Biochimica Et Biophysica Acta-Proteins and Proteomics 1649: 127-139.
5. Zhou HX, Rivas GN, Minton AP (2008) Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. Annual Review of Biophysics. pp. 375-397.
6. Elcock AH (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. Current Opinion in Structural Biology 20: 196-206.
7. Zhou HX (2013) Influence of crowded cellular environments on protein folding, binding, and oligomerization: Biological consequences and potentials of atomistic modeling. FEBS Letters 587: 1053-1061.
8. Phillip Y, Schreiber G (2013) Formation of protein complexes in crowded environments - From in vitro to in vivo. FEBS Letters 587: 1046-1052.
9. Minton AP (2005) Influence of macromolecular crowding upon the stability and state of association of proteins: Predictions and observations. Journal of Pharmaceutical Sciences 94: 1668-1675.
10. Batra J, Xu K, Qin SB, Zhou HX (2009) Effect of macromolecular crowding on protein binding stability: Modest stabilization and significant biological consequences. Biophysical Journal 97: 906-911.
11. Miklos AC, Li CG, Sharaf NG, Pielak GJ (2010) Volume exclusion and soft interaction effects on protein stability under crowded conditions. Biochemistry 49: 6984-6991.
12. Christiansen A, Wang Q, Samiotakis A, Cheung MS, Wittung-Stafshede P (2010) Factors defining effects of macromolecular crowding on protein stability: An in vitro/in silico case study using cytochrome c. Biochemistry 49: 6519-6530.
13. Sanfelice D, Politou A, Martin SR, De Los Rios P, Temussi P, et al. (2013) The effect of crowding and confinement: A comparison of Yfh1 stability in different environments. Physical Biology 10: 7.
14. Hong JA, Gierasch LM (2010) Macromolecular crowding remodels the energy landscape of a protein by favoring a more compact unfolded state. Journal of the American Chemical Society 132: 10445-10452.

15. Mikaelsson T, Aden J, Johansson LBA, Wittung-Stafshede P (2013) Direct observation of protein unfolded state compaction in the presence of macromolecular crowding. Biophysical Journal 104: 694-704.
16. Phillip Y, Sherman E, Haran G, Schreiber G (2009) Common crowding agents have only a small effect on protein-protein interactions. Biophysical Journal 97: 875-885.
17. Miklos AC, Sarkar M, Wang YQ, Pieak GJ Protein crowding tunes protein stability. Journal of the American Chemical Society 133: 7116-7120.
18. Wang YQ, Sarkar M, Smith AE, Krois AS, Pielak GJ (2012) Macromolecular crowding and protein stability. Journal of the American Chemical Society 134: 16614-16618.
19. Ma Q, Fan JB, Zhou Z, Zhou BR, Meng SR, et al. (2012) The contrasting effect of macromolecular crowding on amyloid fibril formation. Plos One 7.
20. Dhar A, Samiotakis A, Ebbinghaus S, Nienhaus L, Homouz D, et al. (2010) Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. Proceedings of the National Academy of Sciences of the United States of America 107: 17586-17591.
21. Heddi B, Phan AT (2011) Structure of human telomeric DNA in crowded solution. Journal of the American Chemical Society 133: 9824-9833.
22. Xue Y, Kan ZY, Wang Q, Yao Y, Liu J, et al. (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K+ solution under molecular crowding condition. Journal of the American Chemical Society 129: 11185-11191.
23. Zheng KW, Chen Z, Hao YH, Tan Z (2010) Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. Nucleic Acids Research 38: 327-338.
24. Homouz D, Sanabria H, Waxham MN, Cheung MS (2009) Modulation of calmodulin plasticity by the effect of macromolecular crowding. Journal of Molecular Biology 391: 933-943.
25. Homouz D, Perham M, Samiotakis A, Cheung MS, Wittung-Stafshede P (2008) Crowded, cell-like environment induces shape changes in aspherical protein. Proceedings of the National Academy of Sciences of the United States of America 105: 11754-11759.
26. Li W, Zhang M, Zhang JL, Li HQ, Zhang XC, et al. (2006) Interactions of daidzin with intramolecular G-quadruplex. FEBS Letters 580: 4905-4910.
27. Chen Z, Zheng KW, Hao YH, Tan Z (2009) Reduced or diminished stabilization of the telomere G-quadruplex and inhibition of telomerase by small chemical ligands under molecular crowding condition. Journal of the American Chemical Society 131: 10430-10438.
28. Rincon V, Bocanegra R, Rodriguez-Huete A, Rivas G, Mateu MG (2011) Effects of macromolecular crowding on the inhibition of virus assembly and virus-cell receptor recognition. Biophysical Journal 100: 738-746.
29. Phillip Y, Harel M, Khait R, Qin SB, Zhou HX, et al. Contrasting factors on the kinetic path to protein complex formation diminish the effects of crowding agents. Biophysical Journal 103: 1011-1019.
30. Wang YQ, Li CG, Pielak GJ Effects of proteins on protein diffusion. Journal of the American Chemical Society 132: 9392-9397.
31. Homouz D, Stagg L, Wittung-Stafshede P, Cheung MS (2009) Macromolecular Crowding Modulates Folding Mechanism of alpha/beta Protein Apoflavodoxin. Biophysical Journal 96: 671-680.

32. Chen E, Christiansen A, Wang Q, Cheung MS, Kliger DS, et al. (2012) Effects of Macromolecular Crowding on Burst Phase Kinetics of Cytochrome c Folding. Biochemistry 51: 9836-9845.
33. Kim YC, Best RB, Mittal J (2010) Macromolecular crowding effects on protein-protein binding affinity and specificity. Journal of Chemical Physics 133: 7.
34. Cheung MS, Klimov D, Thirumalai D (2005) Molecular crowding enhances native state stability and refolding rates of globular proteins. Proceedings of the National Academy of Sciences of the United States of America 102: 4753-4758.
35. Qin SB, Zhou HX (2009) Atomistic Modeling of Macromolecular Crowding Predicts Modest Increases in Protein Folding and Binding Stability. Biophysical Journal 97: 12-19.
36. Mittal J, Best RB (2010) Dependence of Protein Folding Stability and Dynamics on the Density and Composition of Macromolecular Crowders. Biophysical Journal 98: 315-320.
37. Feig M, Sugita Y (2011) Variable Interactions between Protein Crowders and Biomolecular Solutes Are Important in Understanding Cellular Crowding. Journal of Physical Chemistry B 116: 599-605.
38. Harada R, Tochio N, Kigawa T, Sugita Y, Feig M (2013) Reduced Native State Stability in Crowded Cellular Environment Due to Protein-Protein Interactions. Journal of the American Chemical Society 135: 3696-3701.
39. Rosen J, Kim YC, Mittal J (2011) Modest Protein-Crowder Attractive Interactions Can Counteract Enhancement of Protein Association by Intermolecular Excluded Volume Interactions. Journal of Physical Chemistry B 115: 2683-2689.
40. McGuffee SR, Elcock AH (2010) Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. PLoS Computational Biology 6.
41. Abriata LA, Spiga E, Dal Peraro M (2013) All-atom simulations of crowding effects on ubiquitin dynamics. Physical Biology 10: 8.
42. Kim JS, Yethiraj A (2009) Effect of Macromolecular Crowding on Reaction Rates: A Computational and Theoretical Study. Biophysical Journal 96: 1333-1340.
43. Wieczorek G, Zielenkiewicz P (2008) Influence of Macromolecular Crowding on Protein-Protein Association Rates-a Brownian Dynamics Study. Biophysical Journal 95: 5030-5036.
44. Verma PK, Rakshit S, Mitra RK, Pal SK (2011) Role of hydration on the functionality of a proteolytic enzyme alpha-chymotrypsin under crowded environment. Biochimie 93: 1424-1433.
45. Hassan SA, Steinbach PJ (2011) Water-Exclusion and Liquid-Structure Forces in Implicit Solvation. Journal of Physical Chemistry B 115: 14668-14682.
46. Tanizaki S, Clifford J, Connelly BD, Feig M (2008) Conformational sampling of peptides in cellular environments. Biophysical Journal 94: 747-759.
47. Harada R, Sugita Y, Feig M (2012) Protein Crowding Affects Hydration Structure and Dynamics. Journal of the American Chemical Society 134: 4842-4849.
48. Predeus AV, Gul S, Gopal SM, Feig M (2012) Conformational Sampling of Peptides in the Presence of Protein Crowders from AA/CG-Multiscale Simulations. Journal of Physical Chemistry B 116: 8610-8620.
49. Despa F, Fernandez A, Berry RS (2004) Dielectric modulation of biological water. Physical Review Letters 93.

50. Despa F, Fernandez A, Berry RS (2004) Dielectric modulation of biological water (vol 93, art no 228104, 2004). Physical Review Letters 93.
51. Sasmal DK, Ghosh S, Das AK, Bhattacharyya K (2013) Solvation Dynamics of Biological Water in a Single Live Cell under a Confocal Microscope. Langmuir 29: 2289-2298.
52. Tjong H, Zhou HX (2008) Prediction of protein solubility from calculation of transfer free energy. Biophysical Journal 95: 2601-2609.
53. Ghosh S, Chattoraj S, Mondal T, Bhattacharyya K (2013) Dynamics in Cytoplasm, Nucleus, and Lipid Droplet of a Live CHO Cell: Time-Resolved Confocal Microscopy. Langmuir in press.
54. Lee LP, Tidor B (2001) Barstar is electrostatically optimized for tight binding to barnase. Nature Structural Biology 8: 73-76.
55. Lee LP, Tidor B (2001) Optimization of binding electrostatics: Charge complementarity in the barnase-barstar protein complex. Protein Science 10: 362-377.
56. Buckle AM, Schreiber G, Fersht AR (1994) Protein-protein recognition - Crystal structural-analysis of a barnase barstar complex at 2.0-Angstrom resolution. Biochemistry 33: 8878-8889.
57. Brunger AT, Karplus M (1988) Polar hydrogen positions in proteins - Empirical energy placement and neutron-diffraction comparison. Proteins-Structure Function and Genetics 4: 148-156.
58. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) CHARMM - A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry 4: 187-217.
59. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. Journal of Physical Chemistry B 102: 3586-3616.
60. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics 79: 926-935.
61. Altman MD (2006) Computational ligand design and analysis in protein complexes using inverse methods, combinatorial search, and accurate solvation modeling [Ph.D. Thesis]. Cambridge: Massachusetts Institute of Technology.
62. Sitkoff D, Sharp KA, Honig B (1994) Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. Journal of Physical Chemistry 98: 1978-1988.
63. Radhakrishnan ML (2012) Designing electrostatic interactions in biological systems via charge optimization or combinatorial approaches: insights and challenges with a continuum electrostatic framework. Theoretical Chemistry Accounts 131.
64. Hendsch ZS, Tidor B (1999) Electrostatic interactions in the GCN4 leucine zipper: Substantial contributions arise from intramolecular interactions enhanced on binding. Protein Science 8: 1381-1392.
65. Midelfort KS, Hernandez HH, Lippow SM, Tidor B, Drennan CL, et al. (2004) Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. Journal of Molecular Biology 343: 685-701.
66. Green DF, Tidor B (2005) Design of improved protein inhibitors of HIV-1 cell entry: Optimization of electrostatic interactions at the binding interface. Proteins-Structure Function and Bioinformatics 60: 644-657.

67. Carrascal N, Green DF (2010) Energetic Decomposition with the Generalized-Born and Poisson-Boltzmann Solvent Models: Lessons from Association of G-Protein Components. Journal of Physical Chemistry B 114: 5096-5116.
68. Minkara MS, Davis PH, Radhakrishnan ML (2012) Multiple drugs and multiple targets: An analysis of the electrostatic determinants of binding between non-nucleoside HIV-1 reverse transcriptase inhibitors and variants of HIV-1 RT. Proteins-Structure Function and Bioinformatics 80:573-590.
69. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. Journal of Molecular Graphics 14: 33-38.
70. Sheinerman FB, Honig B (2002) On the role of electrostatic interactions in the design of protein-protein interfaces. Journal of Molecular Biology 318: 161-177.
71. Talley K, Ng C, Shoppell M, Kundrotas P, Alexov E (2008) On the electrostatic component of protein-protein binding free energy. PMC Biophysics 1: 2.
72. Giordano S (2003) Effective medium theory for dispersions of dielectric ellipsoids. Journal of Electrostatics 58: 59-76.
73. Gerstein M, Tsai J, Levitt M (1995) The volume of atoms on the protein surface - calculated from simulation, using Voronoi polyhedra. Journal of Molecular Biology 249: 955-966.
74. Bhat S, Purisima EO (2006) Molecular surface generation using a variable-radius solvent probe. Proteins-Structure Function and Bioinformatics 62: 244-261.
75. Li AJ, Nussinov R (1998) A set of van der Waals and Coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. Proteins-Structure Function and Genetics 32: 111-127.
76. Cortines JR, Weigele PR, Gilcrease EB, Casjens SR, Teschke CM (2011) Decoding bacteriophage P22 assembly: Identification of two charged residues in scaffolding protein responsible for coat protein interaction. Virology 421: 1-11.
77. Brown CJ, Srinivasan D, Jun LH, Coomber D, Verma CS, et al. (2008) The electrostatic surface of MDM2 modulates the specificity of its interaction with phosphorylated and unphosphorylated p53 peptides. Cell Cycle 7: 608-610.
78. Busby B, Oashi T, Willis CD, Ackermann MA, Kontrogianni-Konstantopoulos A, et al. (2011) Electrostatic Interactions Mediate Binding of Obscurin to Small Ankyrin 1: Biochemical and Molecular Modeling Studies. Journal of Molecular Biology 408: 321-334.
79. Hu ZJ, Ma BY, Wolfson H, Nussinov R (2000) Conservation of polar residues as hot spots at protein interfaces. Proteins-Structure Function and Genetics 39: 331-342.
80. Qin SB, Zhou HX (2013) FFT-Based Method for Modeling Protein Folding and Binding under Crowding: Benchmarking on Ellipsoidal and All-Atom Crowders. Journal of Chemical Theory and Computation 9: 4633-4643.
81. Christ CD, Mark AE, van Gunsteren WF (2010) Feature Article Basic Ingredients of Free Energy Calculations: A Review. Journal of Computational Chemistry 31: 1569-1582.
82. de Ruiter A, Oostenbrink C (2011) Free energy calculations of protein-ligand interactions. Current Opinion in Chemical Biology 15: 547-552.

Tables

Table I. $\Delta\Delta G_{crowding}$ values for selected model systems described in the text.

$\Delta\Delta G_{crowding}$	LDP	RDP	int	TOT
ε_{out} = 55	-0.1	-1.1	-4.8	-6.0
ε_{out} = 21.9	-3.1	-6.6	-21.7	-31.4
ε_{in} = 4, control run	-1.2 ± 0.5	-1.4 ± 0.5	-1.08 ± 0.05	-3.7 ± 0.7
ε_{in} = 1, same	-3 ± 2	-5 ± 2	-1.4 ± 0.1	-9 ± 3
ε_{in} =1, random	0 ± 1	-5 ± 1	-1.4 ± 0.1	-6 ± 2
0M ions, same, lower grid	-1.2 ± 0.5	-1.4 ± 0.5	-1.08 ± 0.05	-3.7 ± 0.7
0M ions, random, lower grid	-1.0 ± 0.4	-0.1 ± 0.3	-0.97 ± 0.04	-2.1 ± 0.5
0.145M ions, same, lower grid	-0.9 ± 0.5	-1.3 ± 0.5	-0.48 ± 0.05	-2.8 ± 0.7
0.145M ions, random, lower grid	-0.7 ± 0.4	-0.2 ± 0.3	-0.51 ± 0.05	-1.4 ± 0.5

Table I: $\Delta\Delta G_{crowding}$ values broken into components (LDP, RDP, int, and total) for systems not shown in Fig. 3. In the first two rows, the outer dielectric constant is varied as a substitute for explicitly modeling crowders. In the next set of rows (ε_{in} = 1, ε_{in} = 4), the internal dielectric constant was changed to 1 and compared with the control value of the reference system (ε_{in} = 4, also the rightmost bar in Fig. 3). The last four rows show the effect of nonzero ionic strength. For maximal control, all components were re-evaluated at a slightly lowered grid both without ions ("0M ions, same, lower grid") and with ions ("0.145 ions, same, lower grid"). Additionally, crowders were either kept the same as they were in the 50 trials of the reference system ("same") or were randomly varied ("random").

Figures

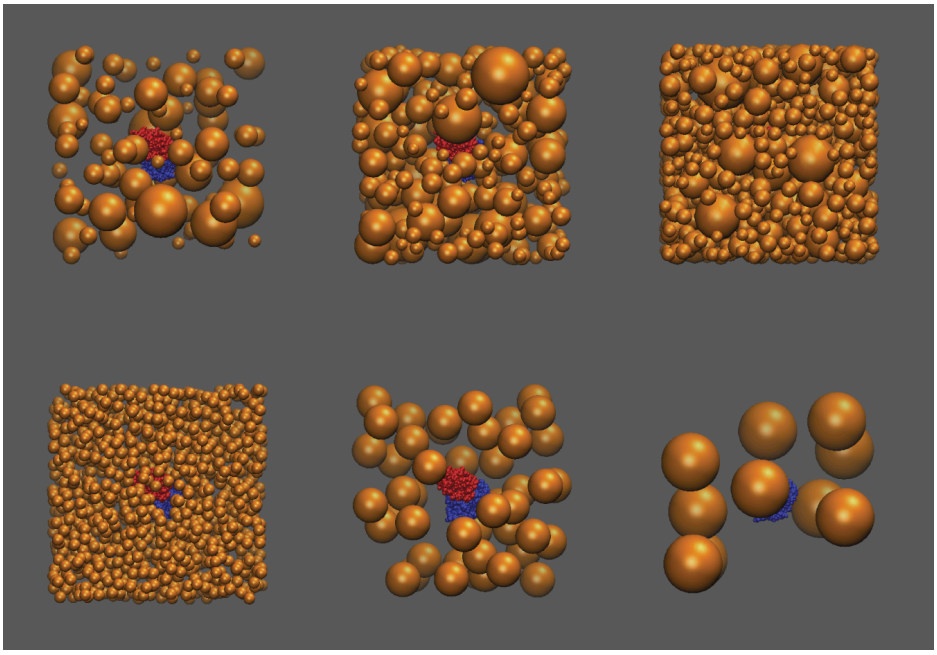


Figure 1. Sample simulated crowded environments. Here, the bound state barnase—barstar complex (red and blue) is surrounded by randomly-placed crowders (orange); the top row depicts environments in which the radius of crowders varied within a system (from 5-25 Å), at increasing crowder volume densities (left to right). The bottom row depicts environments at a constant crowder volume density, but with increasing crowder radius (left to right).

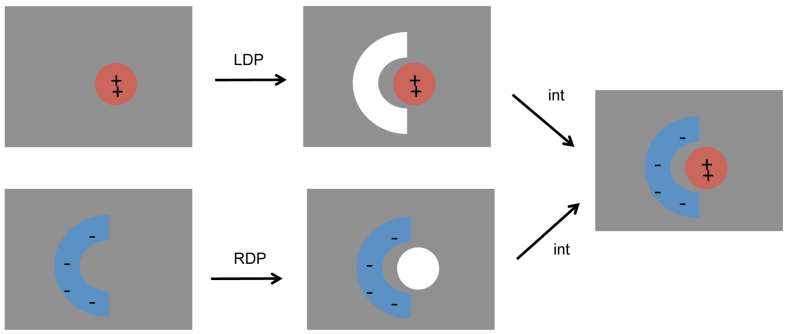


Figure 2: Schematic defining physically relevant components of the electrostatic binding free energy. Pictorially represented are the ligand (barstar) desolvation penalty

(LDP), the receptor (barnase) desolvation penalty, (RDP) and the complex solvent-screened interaction (int). Gray regions denote solvent, and white regions denote low-dielectric cavities in the shape of a given partner, but without charges modeled. The total electrostatic binding free energy is LDP + RDP + int.

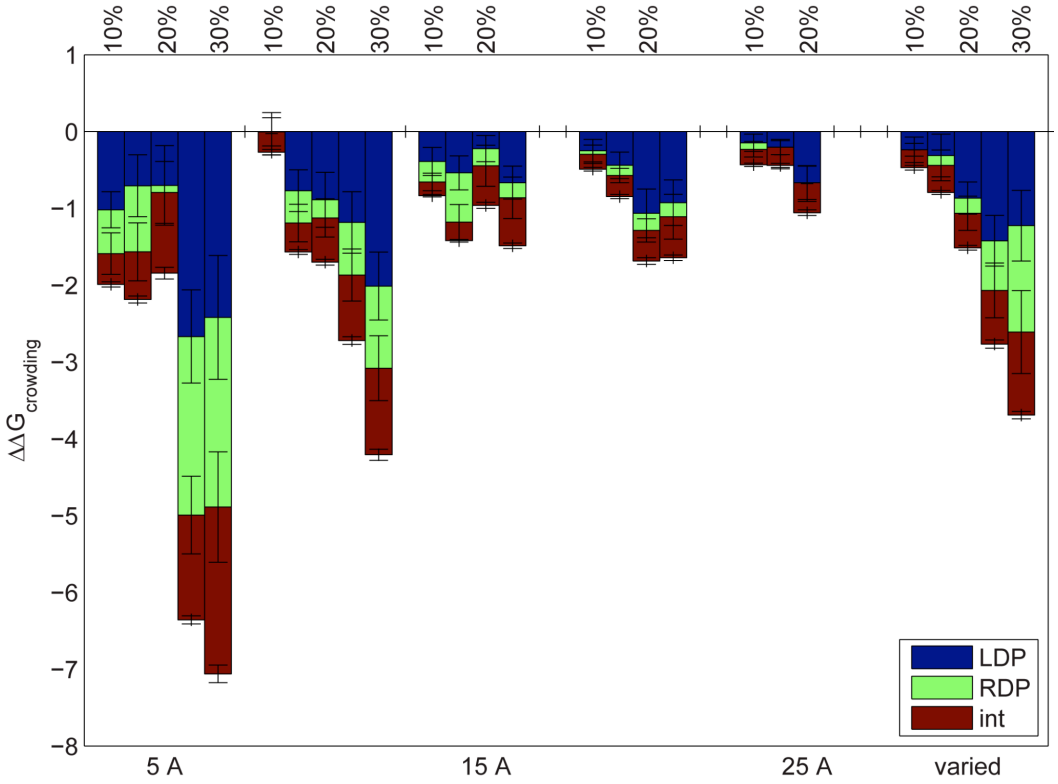


Figure 3: $\Delta\Delta G_{crowding}$, in kcal/mol, for barnase-barstar vs. crowder volume density (top axis) and radius (bottom axis). The bars at right ("varied") are for systems in which the crowder radius varies within each trial. Each bar is the average of 50 trials and is shown as a composite of its contributions of barstar desolvation penalty (LDP, blue), barnase desolvation penalty (RDP, green), and solvent-screened interaction (int, red). Error bars on each contribution represent +/-1 standard error. Missing bars are a result of unsatisfiable geometric constraints (see Results).

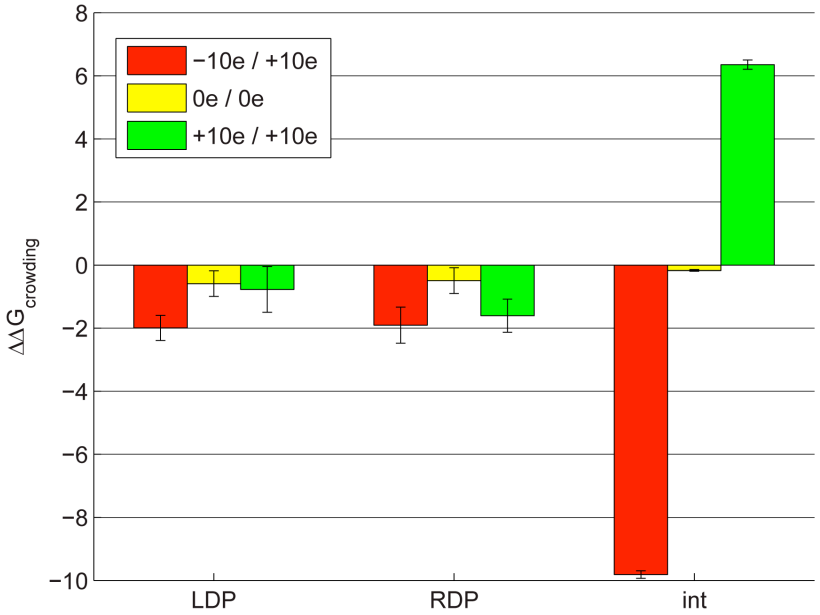


Figure 4: Effect of partner monopole on $\Delta\Delta G_{crowding}$. $\Delta\Delta G_{crowding}$, broken into barstar desolvation penalty (LDP), barnase desolvation penalty (RDP), and solvent-screened interaction (int) components, is shown in kcal/mol for the binding free energy of hypothetical proteins generated by randomly altering the charges of randomly selected atoms on the barnase—barstar complex until a desired overall monopole on each partner is reached (see legend). Each bar shows the average of 50 trials in which the bound and unbound states were crowded with spheres of random, varied radii (5-25 Å) to 30% crowder volume density. Error bars indicate +/-1 standard error.

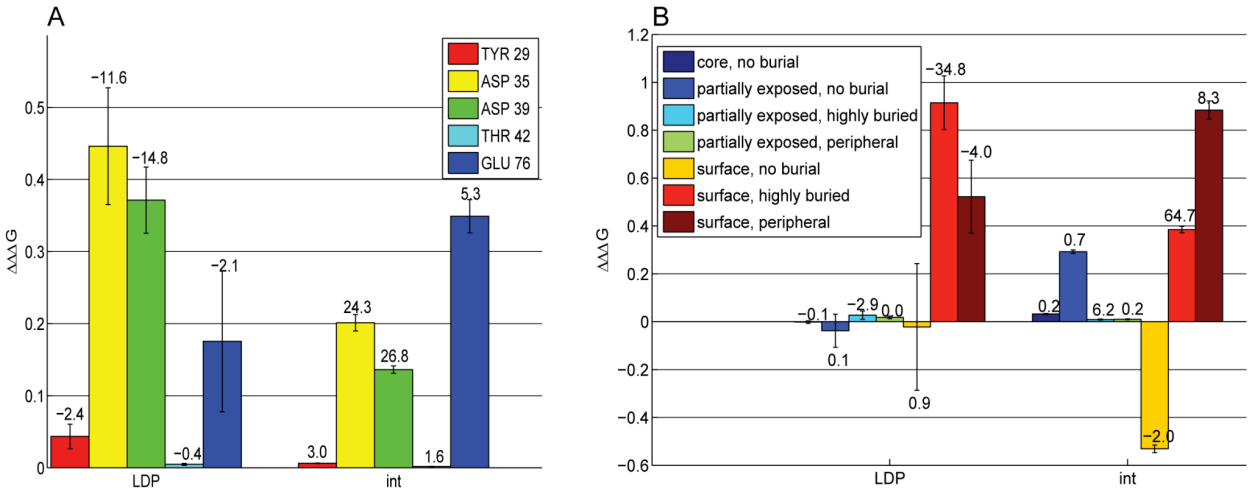


Figure 5: Effect of crowding on residue-based electrostatic contributions.

$\Delta\Delta\Delta G_{res,crowding}$, broken into barstar desolvation penalty (LDP) and interaction (int), in kcal/mol, is shown for (a) selected barstar residues (see legend) and for (b) groups of barstar residues based on level of surface exposure and degree of burial (see Methods); The number above each bar indicates the actual magnitude of the selected component of $\Delta\Delta G_{res}$ without crowding present. Each bar indicates an average of 50 trials in which each crowded bound and unbound state is crowded with spheres of random, varied radii between 5 and 25 Å to 30% crowder volume density. Error bars indicate +/- 1 standard error.

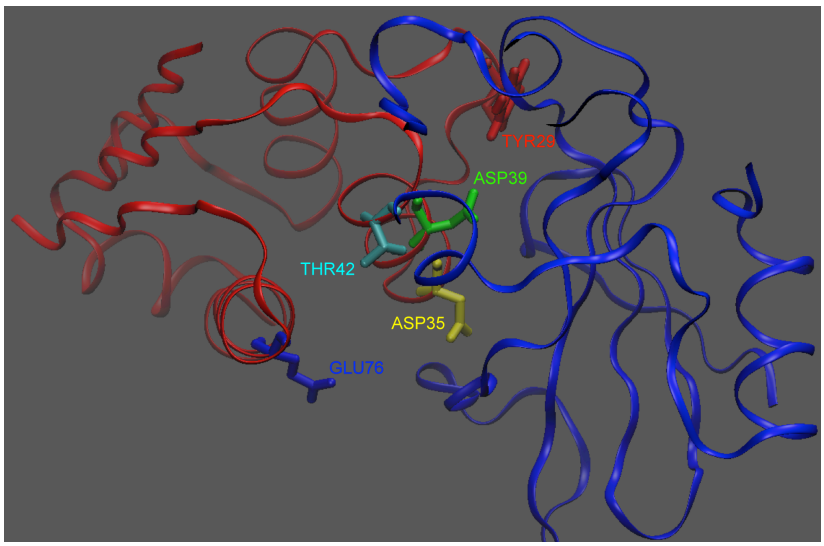


Figure 6: Location of the 5 barstar residues studied via component analysis within the barnase(blue)/barstar(red) complex.

Supporting Information

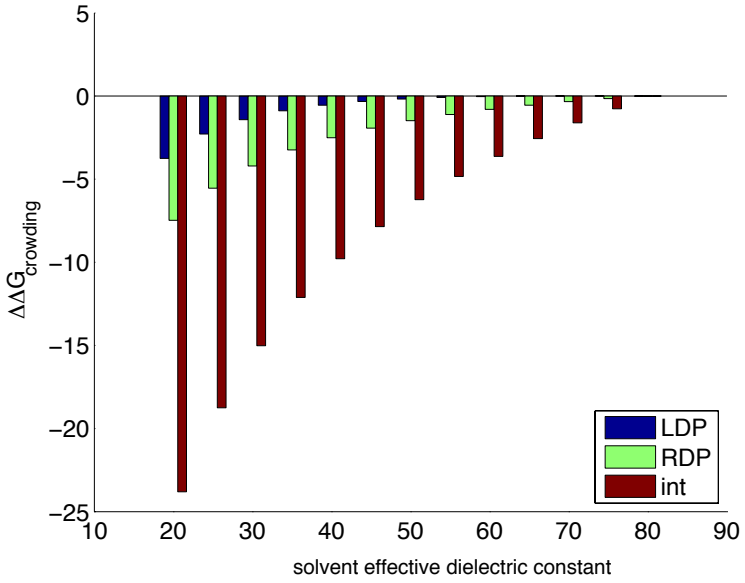


Figure S1: $\Delta\Delta G_{elec}$ vs. solvent dielectric (relative to a solvent dielectric constant of 80), without explicit crowders. A lowering of the external dielectric constant produces a

similar qualitative trend as increasing the volume density or decreasing the radius of explicit crowders.

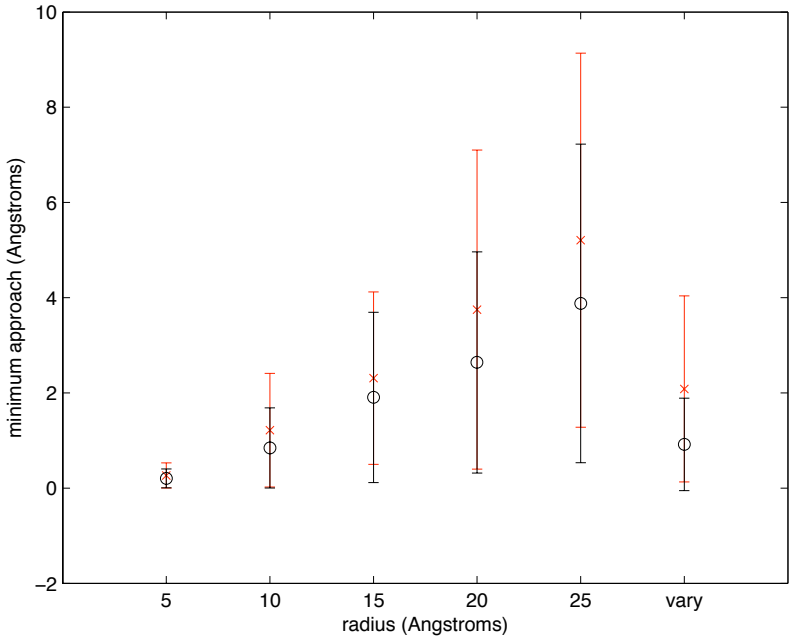


Figure S2: Average minimum distance of approach between crowders and protein vs. crowder radius. The minimum distance of approach is the shortest distance between the protein and crowder in each state, accounting for their radii. Data are shown for both 15% crowder volume density (data for 20% crowder density show a similar trend, not shown). Data are averaged over bound and unbound states for all 50 trials conducted for each radius and volume density. Error bars are +/- one standard deviation.

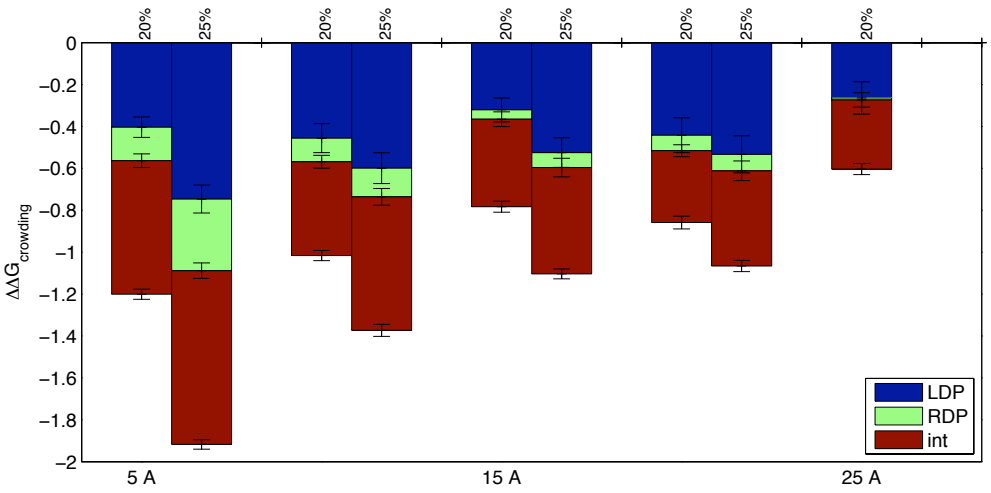


Figure S3: Effect on $\Delta\Delta G_{crowding}$ of using a zero-radius probe to generate the molecular surface. A subset of runs shown in Fig. 3 were redone using a zero-radius probe sphere to generate the molecular surface instead of the standard 1.4-Å probe. Identical crowder placements were used for each bar shown here and the bar corresponding to the same crowder density and radii in Fig. 3; the only different is in the size of the probe sphere.

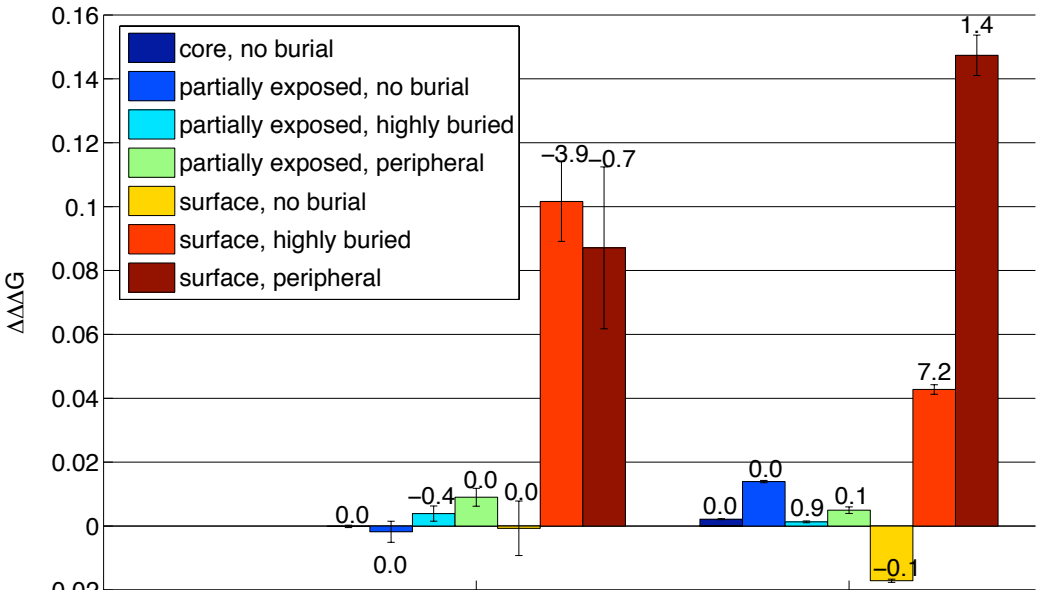


Figure S4. Per residue ΔΔΔG for sets of residues on barstar. Residues were grouped by degree of burial and solvent exposure and values were *normalized by dividing by the number of residues in each group* (Figure 5b in the main text does not normalize per residue). Similar overall qualitative trends are seen in this Figure and in Figure 5b in the main text. The number above each bar indicates the per-residue value of the selected component of $\Delta\Delta G_{res}$