

Separated topologies—A method for relative binding free energy calculations using orientational restraints

Gabriel J. Rocklin,^{1,2,a)} David L. Mobley,^{3,4} and Ken A. Dill⁵

¹Department of Pharmaceutical Chemistry, University of California San Francisco, 1700 4th St., San Francisco, California 94143-2550, USA

²Biophysics Graduate Program, University of California San Francisco, 1700 4th St., San Francisco, California 94143-2550, USA

³Departments of Pharmaceutical Sciences and Chemistry, University of California Irvine, 147 Bison Modular, Building 515, Irvine, California 92697-0001, USA

⁴Department of Chemistry, University of New Orleans, 2000 Lakeshore Drive, New Orleans, Louisiana 70148, USA

⁵Laufer Center for Physical and Quantitative Biology, 5252 Stony Brook University, Stony Brook, New York 11794-0001, USA

(Received 18 December 2012; accepted 31 January 2013; published online 25 February 2013)

Oriental restraints can improve the efficiency of alchemical free energy calculations, but they are not typically applied in relative binding calculations, which compute the affinity difference between two ligands. Here, we describe a new “separated topologies” method, which computes relative binding free energies using orientational restraints and which has several advantages over existing methods. While standard approaches maintain the initial and final ligand in a shared orientation, the separated topologies approach allows the initial and final ligands to have distinct orientations. This avoids a slowly converging reorientation step in the calculation. The separated topologies approach can also be applied to determine the relative free energies of multiple orientations of the same ligand. We illustrate the approach by calculating the relative binding free energies of two compounds to an engineered site in Cytochrome C Peroxidase. © 2013 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4792251>]

INTRODUCTION

Molecular simulations are often used to compute relative protein-ligand binding free energies, i.e., the difference in binding free energy between an initial ligand A and a final ligand B.^{1–3} These calculations typically rely on the thermodynamic cycle in Figure 1, in which A is transformed *alchemically* (i.e., by directly changing simulation parameters) across a series of intermediate simulations into B both in the binding site ($\Delta\Delta G_{\text{site}}$) and in solution ($\Delta\Delta G_{\text{sol}}$). While these two transformations are alchemical and do not represent physically possible processes, their free energies can be used to compute the free energy of the physical process of interest: the difference between the binding free energies of A and B, or $\Delta G^{\circ}_{\text{bind,A}} - \Delta G^{\circ}_{\text{bind,B}}$.

Two main approaches exist for calculating $\Delta\Delta G_{\text{site}}$, the free energy of transforming A into B in the binding site: the *single topology* approach and the *dual topologies* approach. In the single topology approach, there is one ligand molecule in all simulations, which morphs from being in state A (having a particular chemical structure and interaction parameters), to a different state B, where the structure and interaction parameters are different.^{4–13} In some transformations, there could be atoms on A that have no direct analog on B. During the computational morphing process, these atoms will have their non-bonded interactions decoupled from the rest of the sys-

tem, meaning that these atoms (now referred to as *dummy atoms*) on A will remain bonded to the morphing ligand but will no longer interact with the protein or solvent.¹² Similarly, there may be atoms that are unique to B. When transforming A into B, these atoms would begin as dummy atoms and become coupled to the system (protein plus solvent) during the computational morphing process. Despite the ability to employ dummy atoms, the single topology approach is easiest to apply when the atoms and bonds on A map easily to the atoms and bonds on B. As a result, it is commonly used when A and B are near-superimposable or when one ligand is a substructure of the other.

In the dual topologies approach, the computational morphing process carries along two ligand molecules at the same time throughout all simulations.^{14–16} Using this method, the simulations begin with both a “real” ligand A and a “dummy” ligand B, and over the course of the alchemical transformation, A becomes the dummy ligand while B becomes the real ligand. That is, in the initial state the non-bonded interactions of B are fully decoupled from the rest of the system (the protein and the solvent) and in the final state the non-bonded interactions of A are fully decoupled. At both the initial and final endpoints of the transformation, only one ligand is coupled to the system. In the intermediate states, both ligands are partially coupled to the protein and solvent, but neither ligand interacts with the other. The dual topologies approach is convenient because no mapping is required between the atoms and bonds on A and B. However, because the approach involves an entirely decoupled molecule at each

^{a)} Author to whom correspondence should be addressed. Electronic mail: rocklin@gmail.com.

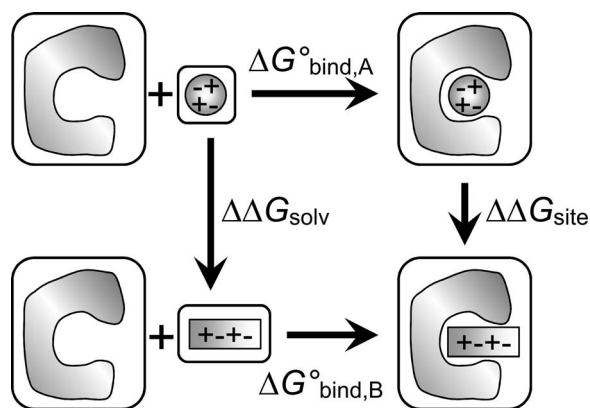


FIG. 1. The common thermodynamic cycle for relative binding free energies. The difference in free energies of the alchemical steps $\Delta\Delta G_{\text{site}} - \Delta\Delta G_{\text{solv}}$ provides the difference in free energies of the physical steps $\Delta G^{\circ}_{\text{bind,A}} - \Delta G^{\circ}_{\text{bind,B}}$.

endpoint (rather than just select decoupled atoms, still bonded to the ligand topology), restraints must be employed to maintain the weakly coupled or fully decoupled ligand in the binding site, in order to avoid the need for it to sample the entire simulation cell volume.

Several forms of this restraint have been previously proposed.

(i) **Holding A and B to each other:** A harmonic distance restraint has been used to connect A to B, so that the decoupled ligand cannot diffuse away from the coupled ligand.¹⁶ Other methods use constraints—non-energetic terms which are enforced at each timestep—to connect A and B, in some instances by performing identical MC moves on both ligands.^{14,15} Restraints which connect A to B do not alter the rotational and translational entropy of either ligand in the fully coupled state, because even though the fully coupled A is always attached to B, B does not interact with the environment and thus does not alter the free energy landscape felt by A.

(ii) **Holding A and B to the binding site:** An alternate approach is to restrain both A and B to the binding site. In the lambda dynamics approach, a restraint potential proportional to lambda is applied to the ligands, such that the fully coupled ligand feels no restraint potential, but the more decoupled ligands feel the restraint more strongly.^{17,18} This restraint takes the form of a so-called *ghost force* in which the binding site non-bonded interactions create forces (of reduced magnitude, dependent on lambda) on the decoupled ligand, but the ligand does not create opposing forces on the binding site, creating non-Newtonian dynamics. In these works, the authors chose not to directly account for the effect of the restraint on the translational and rotational entropy of the ligands, instead arguing that these terms would be similar for different ligands in the same binding site and would not affect their relative binding free energies.

Each of these approaches allows the ligands to orient freely in the binding site during the free energy calculation. In other words, the ligands are placed into the binding site in particular (known or guessed) orientations, and then if these orientations are not the global minimum orientations for either ligand, that ligand must reorient to obtain a converged

result.¹⁹ This reorientation will often be the slowest factor in obtaining a converged result for the free energy difference. If the ligands are orientationally restrained to each other (as a result of, for example, a single topology, but this also applies to the method of Michel *et al.*¹⁴), both alchemical endpoint simulations need to find their global minimum orientations, and the alchemical intermediate simulations must sample back and forth between both orientations, which may require the slow crossing of a significant energy barrier. These slow transitions between orientations will be required even if Hamiltonian exchange^{20,21} methods are applied. To further complicate matters, if ligands do need to reorient, all simulation data in this unequilibrated “orientational sampling” phase must be discarded from the analysis prior to calculating the free energy differences. This unequilibrated data can have a significant effect on the free energy estimates, unless the simulations are extended to such a length that all ligand orientations have been sampled in their equilibrium proportions, which requires a sampling length covering many transitions between orientations, rather than just one transition to a ligand’s global energy minimum.

THE SEPARATED TOPOLOGIES METHOD

Here, we propose a new method for computing relative binding free energies, which is a variant of the dual topologies method. We call this approach the method of separated topologies. Unlike in the dual topologies method, the two ligands are orientationally restrained to the binding site in their decoupled states, rather than being restrained or constrained to each other. These orientational restraints are commonly applied in absolute binding free energy calculations.^{22,23,12,24} In essence, two absolute binding calculations are performed at once, in different directions, and overlapping in the middle. The path (representing transformation $\Delta\Delta G_{\text{site}}$ in Figure 1) is shown in Figure 2. In the initial state, A is unrestrained, fully coupled, and sampling a particular orientation in the protein and B is fully decoupled from the protein in an ideal gas at 1 M standard concentration. The process is described step by step below, with comments afterwards.

- (1) A is restrained in the binding site, using a distance, two angles, and three dihedral restraints. Unsampled symmetric orientations of A are included through a symmetry correction.
- (2) B is restrained from 1 M volume to the binding site, using its own set of six total distances, angles, and dihedral restraints.
- (3) The Van der Waals (VdW) interactions of B are coupled to the protein (using soft-core interactions), though B is non-interacting with A.
- (4) The partial charges on A are decoupled from the protein, while those of B are coupled in.
- (5) The VdW interactions of A are decoupled from the protein (again using soft-core interactions).
- (6) The restraints on A are released, moving A from its restrained volume to the 1 M standard state volume.
- (7) The restraints on B are released, eliminating any non-physical forces in the simulation. Unsampled symmetric

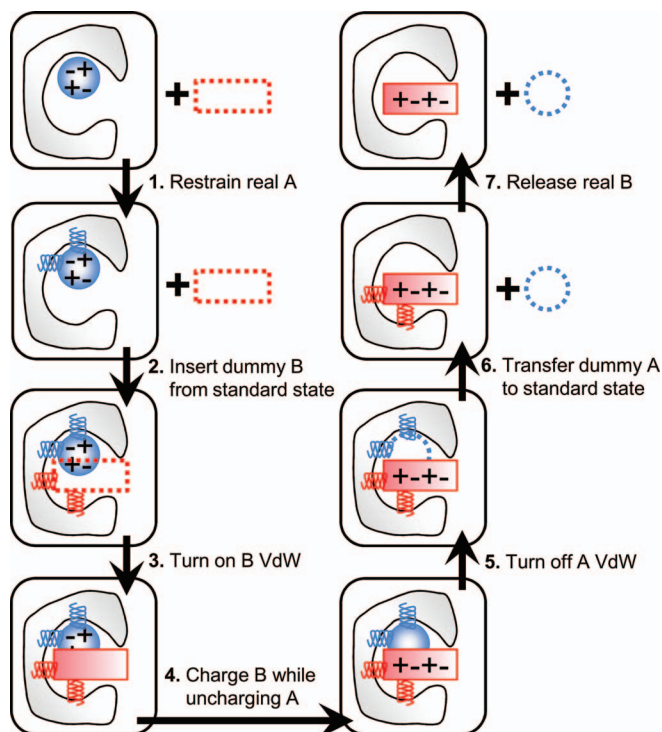


FIG. 2. Separated topologies method for calculating $\Delta\Delta G_{\text{site}}$. Initially, ligand A (blue circle) is bound to the protein, while ligand B (red rectangle) is at standard 1 M concentration in the gas phase (indicated by an empty rectangle outside of the simulation cell). First, ligand A is restrained to the protein ($\Delta G_{\text{rest,A}}$), indicated by adding springs. Ligand B is also then restrained to the protein ($\Delta G_{\text{ss,B}}$) and this free energy is computed analytically. The Lennard-Jones interactions of ligand B are then coupled to the protein ($\Delta G_{\text{vdw,B}}$), filling the rectangle. Next, the partial charges on ligand A are turned to zero simultaneously with the partial charges on ligand B being introduced (ΔG_{chg}). The Lennard-Jones interactions on ligand A are then decoupled from the protein ($\Delta G_{\text{vdw,A}}$), and then the restraints on ligand A are released (computed analytically) so that ligand A moves to the gas phase at 1 M volume ($\Delta G_{\text{ss,A}}$). Finally, the restraints on ligand B are released ($\Delta G_{\text{rest,B}}$), leaving a fully coupled ligand B alone in the binding site with no remaining unphysical restraining potentials. Note that the transformations on the right are oppositely signed from the transformations on the left: while ($\Delta G_{\text{rest,A}}$), consists of applying harmonic restraints on A, ($\Delta G_{\text{rest,B}}$) consists of releasing pre-existing harmonic restraints on B, and likewise for the standard (ss) and VdW transformations.

orientations of B are included through a symmetry correction.

In the final state, symmetric with the initial state, B is unrestrained, fully coupled to the binding site, and sampling a particular orientation, while A is fully decoupled from the protein in an ideal gas at 1 M standard concentration. The restraints serve as computational devices to maintain the ligands in a particular orientation, but do not affect the overall relative binding free energy because they are released for each ligand in that ligand's bound, fully coupled state and each ligand is at a standard 1 M concentration in its fully decoupled state. The bound state for each ligand is defined implicitly as the configurations which it samples when unrestrained, which should be entirely in the binding site for typical simulation lengths. One can also modify the central steps of the path, performing these alternate steps (3b), (4b), and (5b) instead

(3b) The partial charges on A are decoupled from the protein.

(4b) The VdW interactions of A are decoupled from the protein, while those of B are coupled in.

(5b) The partial charges on B are coupled to the protein.

The original cycle will be preferable in cases where A and B each have (the same) net charge, because the net charge of the system will then be conserved throughout the whole transformation, which improves convergence. Even with neutral but polar ligands, the original cycle is likely to be preferable because, assuming A and B both make similar polar interactions with the protein, these polar interactions will be maintained throughout the transformation, rather than the artificial creation of a completely nonpolar molecule in a polar binding site during steps (3b) and (5b). If soft-core charge perturbations are applied in addition to soft-core VdW perturbations,¹⁴ the charge and VdW parameters could be perturbed simultaneously and steps (3)–(5) could be combined into a single step. The overall goal in choosing intermediates for steps (3)–(5) should be to minimize the reorganization of the protein and solvent between different states. This is why we suggest maintaining a fully VdW-coupled ligand in the binding site at all times during the transformation and the same net charge (if possible) in the binding site at all times.

The free energies of steps (2) and (6) are computed analytically, using the formula in Boresch *et al.*²² (2003). The free energies of steps (3), (4), and (5) can be computed by any standard method such as the Zwanzig equation, thermodynamic integration (TI), the Bennett acceptance ratio (BAR),²⁵ or the multistate Bennett acceptance ratio (MBAR).²⁶ Finally, to compute the free energies of steps (1) and (7) efficiently, the orientational restraints on the ligands should be set such that the energy minimum of the restraints corresponds to the pre-existing energy minimum for the ligand's orientation in the protein environment, and simulations should be started with the ligand in this orientation. When restraints are used in this manner, the free energies of steps (1) and (7) converge very rapidly because the phase space of the restrained state is simply the phase space of the unrestrained state, biased even more strongly toward its most common configurations. A single simulation of the unrestrained state (analyzed using the Zwanzig equation) is typically adequate, although because a simulation of each fully coupled, restrained state is already performed as the initial intermediate in step (3) or the final intermediate in step (5), BAR can be easily applied as well. Fast convergence of these steps also relies on the ligand only sampling a single orientation (a single energy well or metastable state). After discussing the disadvantages and advantages of this approach compared with existing methods for computing relative free energies, we will discuss the case where the binding of A or B cannot be described exclusively by a single orientation.

Like the dual topologies approach, the separated topologies approach calculates $\Delta\Delta G_{\text{solv}}$ by performing vacuum-to-water transfer (hydration) free energy calculations on both ligand A and ligand B, and taking the difference. This is different from the single topology approach, in which $\Delta\Delta G_{\text{solv}}$ is calculated by transforming the atoms and bonds on A into the atoms and bonds on B in solution. This single topology approach to $\Delta\Delta G_{\text{solv}}$ will normally include the

free energy of changing the bonded and internal (ligand atom to other ligand atom) charge and/or VdW interactions as part of $\Delta\Delta G_{\text{solv}}$. This is acceptable because these internal interactions are also included in the single topology calculation for $\Delta\Delta G_{\text{site}}$. However, these internal interactions are *not* included in the separated or dual topology calculation for $\Delta\Delta G_{\text{site}}$, which means they cannot be included either in the separated or dual topologies calculation for $\Delta\Delta G_{\text{solv}}$. When using the separated topologies method for $\Delta\Delta G_{\text{site}}$, calculating $\Delta\Delta G_{\text{solv}}$ using relative hydration free energies will ensure that these internal energetic terms are treated consistently between $\Delta\Delta G_{\text{solv}}$ and $\Delta\Delta G_{\text{site}}$.

The two main disadvantages of separated topologies are (1) two VdW steps (3 and 5) or two charging steps (3b and 5b) are required instead of one and (2) an additional restraining step must be performed for each ligand (steps 2 and 6). However, the first disadvantage is inevitable to any approach in cases where dummy atoms must be used on both ligands. The second disadvantage can be strongly mitigated by choosing restraints that match the pre-existing energy minimum, so that the restraining energies converge very rapidly.

The first advantage of separated topologies is that A and B are not required to share an orientation. Even very similar ligands in simple binding sites often adopt different orientations and the kinetic barriers between orientations can be significant enough that ligands will not reorient during relative free energy calculations using the standard approaches reviewed earlier (cf. the difference between relative transformations with phenol and catechol in Boyce *et al.* (2009)¹²). Also, even if ligands eventually reorient, this step can be rate limiting to converging the relative binding free energy, meaning that using a cycle which does not require reorientation will be more efficient.²⁷ As free energy calculations are applied to more complex ligands, these reorientations will become more significant and more costly. Some dual topology methods do allow each ligand to be initially positioned in different orientations (and held together with, for instance, a single distance restraint). Still, these methods allow each ligand to reorient while it is decoupled. This increases the amount of sampling necessary in the decoupled state, compared with the separated topologies approach, which restrains each ligand to the binding site in a single orientation when that ligand is decoupled. Finally, the separated topologies approach maintains the advantages of relative binding calculations over computing relative binding through absolute binding calculations: it is likely to be much easier to replace any ligand with any other ligand in a binding site than it is to replace either ligand with solvent and this is especially true in the case of ligands with net charge.

The second advantage of separated topologies is that it provides a clear and consistent means for explicitly sampling multiple orientations of each ligand. In some cases, multiple metastable orientations may contribute significantly to the binding affinity of a ligand or one may not know which orientation is dominant in advance of the calculation, making it impossible to set the orientational restraints to match the global minimum orientation *a priori*. In these cases, we propose performing a separate calculation on each pose of the ligand, as described in Mobley *et al.*²⁷ for absolute calculations. For ex-

ample, if ligand A has two metastable orientations A1 and A2, two relative free energy calculations can be performed using the dual topologies cycle described above. In the first calculation, A1 is transformed into B. In the second calculation, A1 is transformed to A2. For this second calculation (A1 into A2), the free energy of step (1) represents restraining A from exclusively the A1 phase space to a set of orientational restraints that match the global minimum of A1. The free energy of step (7), in this instance, represents restraining “B” (which is really a second copy of A) from exclusively the A2 phase space to orientational restraints that match A2. Dividing the phase space in this manner requires an indicator function, so that all snapshots can be classified as A1 or A2 phase space. In Mobley *et al.*,²⁷ a single degree of freedom was used as the indicator function to determine if a given configuration of A was a member of A1 or A2. We propose instead an energy-based indicator: any configuration of A that restrains more easily to the A1 restraints (compared to the A2 restraints) is a member of the A1 phase space and any configuration of A that restrains more easily to the A2 restraints is a member of the A2 phase space. Thus defining the restraints automatically defines the indicator function.

When combined with the relative free energies of hydration, these two relative calculations (A1 \rightarrow B and A1 \rightarrow A2) can be used to compute the overall relative binding free energy between A and B. One first computes the relative free energy of binding of A (i.e., A1 or A2) relative to A1, using a modified form of Eq. (12) in Mobley *et al.* (2006)²⁷

$$\Delta\Delta G^\circ = -\beta^{-1}\ln[\exp(-\beta^{-1}\Delta\Delta G_1^\circ) + \exp(-\beta^{-1}\Delta\Delta G_2^\circ)], \quad (1)$$

where $\Delta\Delta G^\circ$ is the difference in binding free energy between A1 and A, $\Delta\Delta G_1^\circ$ is the difference in binding free energy between A1 and A1 (i.e., 0), $\Delta\Delta G_2^\circ$ is the difference in binding free energy between A1 and A2 (from the second alchemical calculation), and $\beta = (k_B T)^{-1}$, where k_B is the Boltzmann constant and T is the absolute temperature. Once the relative binding free energy between A and A1 has been calculated as above, A and B can be directly compared via the comparison between A1 and B.

In the above protocol, the separated topologies method was applied to compute both the relative free energies of A1 and B, and A1 and A2. The method is particularly advantageous at comparing orientations of the same ligand for two reasons. First, the method takes an alchemical route between the different orientations, which means that the fully coupled ligand never needs to sample the high-energy, physical transition state(s) between the orientations, as in umbrella sampling. Second, any method for computing relative free energies of orientations requires an indicator function for defining the orientations. In the separated topologies method, the indicator function follows naturally from the restraints, because each set of restraints defines an energy well around a particular orientation. These restraints can be defined automatically based on unrestrained simulations of the ligand in a given orientation and make it unnecessary to define indicator functions by hand or compute RMSD's for use as an indicator function.

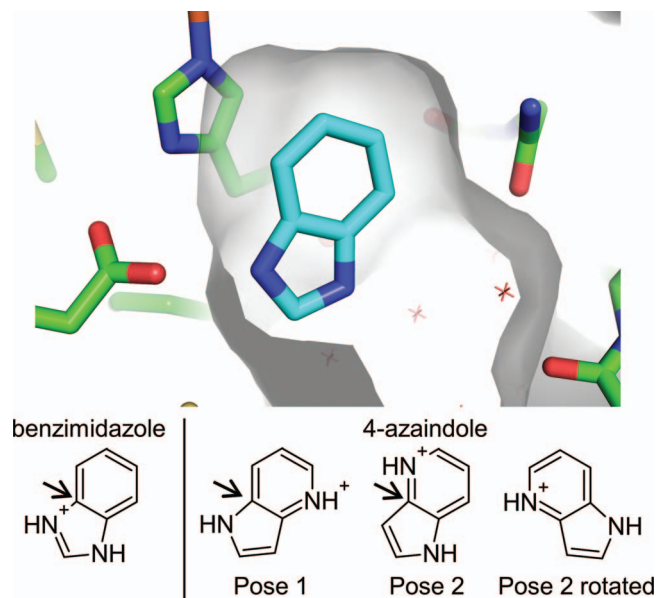


FIG. 3. (top) Benzimidazole bound to CCP-GA, from Ref. 22. Figure generated using PyMOL 1.4.1, Schrödinger, LLC. (bottom) Structures of benzimidazole and 4-azaindole, showing the two ways of superimposing 4-azaindole on benzimidazole. When unrestrained in Pose 2, 4-azaindole rapidly rotates to the more stable “Pose 2 rotated” orientation in order to hydrogen bond to the protein Asp. Arrows indicate the attachment point for the restraint using the dual topologies method.

RESULTS

We tested three different methods to calculate the relative binding free energy of benzimidazole (ligand A) and 4-azaindole (ligand B) to Cytochrome C Peroxidase W191G “Gateless” (CCP-GA)²⁸ (Figure 3). This test had two main objectives. First, we show that our method (Figure 2) provides the same free energy estimate as two existing approaches for calculating relative free energies, in the case where ligand reorientation is not needed. Second, we show that when ligand reorientation is required, existing approaches will fail to converge, but our separated topologies method still produces the correct answer. The three tested approaches are described briefly below and details of the simulations are given in Methods.

Method 1: Single topology

To calculate $\Delta\Delta G_{\text{site}}$ and $\Delta\Delta G_{\text{solv}}$, we transformed the atoms and bonds in benzimidazole into the atoms and bonds of 4-azaindole in the binding site and in solution.

Method 2: Dual topologies

In this approach, individual topologies for both benzimidazole and 4-azaindole exist in the binding site, attached to each other using a $10 \text{ kcal/mol/\text{Å}^2}$ harmonic restraint on the atoms indicated by an arrow in Figure 3. Neither ligand interacted with the other through non-bonded forces. To calculate $\Delta\Delta G_{\text{site}}$, we transformed benzimidazole into 4-azaindole by coupling the VdW interactions of 4-azaindole to the protein ($\Delta G_{\text{vdw,B}}$), followed by uncharging benzimidazole while charging 4-azaindole (ΔG_{charge}), followed by decou-

pling the VdW interactions of benzimidazole from the protein ($\Delta G_{\text{vdw,B}}$). To calculate $\Delta\Delta G_{\text{solv}}$, we calculated the relative hydration free energy of the two compounds.

Method 3: Separated topologies

In this approach, we used the proposed method in the Introduction (Figure 2) to calculate $\Delta\Delta G_{\text{site}}$. To calculate $\Delta\Delta G_{\text{solv}}$, we again used the relative hydration free energy of the two compounds.

Benzimidazole and 4-azaindole are superimposable and differ in the location of one nitrogen heteroatom. A relative free energy calculation would typically start by superimposing 4-azaindole on benzimidazole, but because of the symmetry of benzimidazole, there are two possible orientations for 4-azaindole. In *Pose 1*, the indole nitrogen of 4-azaindole (N1) is superimposed with the nitrogen on benzimidazole which hydrogen bonds to the protein aspartate. In *Pose 2*, N1 of 4-azaindole is superimposed on the opposite nitrogen of benzimidazole and initially makes no hydrogen bond with the aspartate. When simulated without any restraints applied, this pose rapidly rotates counterclockwise (in the frame of Fig. 3) so that the pyridine nitrogen of 4-azaindole hydrogen bonds with the protein aspartate (Fig. 3 bottom, “Pose 2 rotated”). The ligand then remains in this orientation for the length of our simulations without ever sampling Pose 1. To test how the results of each method depend on the starting orientation of 4-azaindole, we performed the relative free energy calculation using each approach twice, once with 4-azaindole starting in Pose 1 and once starting from Pose 2.

In Table I, we show the results of each method using each starting orientation of 4-azaindole. The convergence of the free energy estimates with simulation time (up to 5 ns per lambda intermediate) is shown in Figure 4. Even with five nanoseconds of simulation time, the overall free energy estimates ($\Delta\Delta G_{\text{bind}}$) using both the single topology method and the dual topology method show a large dependence on the starting orientation of 4-azaindole. Using the single topology method, the simulations started from Pose 2 never sample the Pose 1 orientation. Using the dual topologies method, the simulations started from Pose 2 only sample Pose 1 during simulations in which the VdW interactions of 4-azaindole are partially decoupled. Because neither method, when started from Pose 2, ever sampled the more favorable 4-azaindole orientation (Pose 1) in the fully coupled state, both methods overestimate how weak a binder 4-azaindole is relative to benzimidazole.

Our method of separated topologies gives similar results to both the single topology and dual topologies methods when started with 4-azaindole in either pose (Table I). To compute a formally correct $\Delta\Delta G_{\text{bind}}$ using separated topologies with both the Pose 1 and Pose 2 data, the $\Delta\Delta G_{\text{bind}}$ values from each pose need to be combined according to Eq. (1). Using the values in Table I and 0.596 kcal/mol as $k_B T$, an overall $\Delta\Delta G_{\text{bind}}$ of 2.208 kcal/mol is returned, nearly identical to the value obtained from using the Pose 1 data alone.

It is more difficult to combine the Pose 1 and Pose 2 data into a single affinity when using the single topology method or the dual topologies method. The separated topologies method

TABLE I. Results of relative free energy calculations transforming benzimidazole into 4-azaindole, broken down into individual components. $\Delta\Delta G_{\text{solv}}$ differs between Method 1 and Methods 2 and 3 because changes in the internal bonded and Lennard-Jones energies of the ligands are included in Method 1. Changes in internal electrostatic energies are not included in the quantities for any method. $\Delta\Delta G_{\text{bind}}$ is calculated as $\Delta\Delta G_{\text{site}} - \Delta\Delta G_{\text{solv}}$; $\Delta\Delta G_{\text{site}}$ is the sum of all the above components for Methods 2 and 3. Free energy components for Method 3 are defined in Fig. 2.

Method 1: Single topology		
	Pose 1	Pose 2
$\Delta\Delta G_{\text{site}}$	8.87 ± 0.03	12.33 ± 0.06
$\Delta\Delta G_{\text{solv}}$	7.10 ± 0.04	7.10 ± 0.04
$\Delta\Delta G_{\text{bind}}$	1.78 ± 0.05	5.23 ± 0.08
Method 2: Dual topologies		
	Pose 1	Pose 2
$\Delta G_{\text{vdw},B}$	-18.09 ± 0.04	-18.45 ± 0.04
$\Delta G_{\text{charge}} + \Delta G_{\text{vdw},A}$	20.53 ± 0.05	24.53 ± 0.04
$\Delta\Delta G_{\text{site}}$	2.44 ± 0.06	6.08 ± 0.06
$\Delta\Delta G_{\text{solv}}$	0.67 ± 0.07	0.67 ± 0.07
$\Delta\Delta G_{\text{bind}}$	1.77 ± 0.09	5.41 ± 0.09
Method 3: Separated topologies (Fig. 2)		
	Pose 1	Pose 2
$\Delta G_{\text{rest},A}$	0.89 ± 0.02	0.89 ± 0.02
Symmetry correction on A	0.41	0.41
$\Delta G_{\text{ss},B}$	8.38	8.34
$\Delta G_{\text{vdw},B}$	-20.89 ± 0.02	-20.57 ± 0.02
$\Delta G_{\text{charge}} + \Delta G_{\text{vdw},A}$	23.61 ± 0.03	27.57 ± 0.03
$\Delta G_{\text{ss},A}$	-8.38	-8.38
$\Delta G_{\text{rest},B}$	-1.14 ± 0.16	-2.01 ± 0.10
Symmetry correction on B	0	0
$\Delta\Delta G_{\text{site}}$	2.88 ± 0.16	6.25 ± 0.10
$\Delta\Delta G_{\text{solv}}$	0.67 ± 0.07	0.67 ± 0.07
$\Delta\Delta G_{\text{bind}}(\text{pose})$	2.21 ± 0.17	5.58 ± 0.13
$\Delta\Delta G_{\text{bind}}(\text{total, by Eq. (1)})$	2.21 ± 0.17	

begins by assigning a set of restraints to each pose and these restraints automatically define an indicator function because any configuration of the ligand will restrain most easily to a single set of restraints. The single topology and dual topology methods do not require restraints, though it would still be possible to divide all simulation data into separate poses by defining an indicator function, as some have done.²⁹ This would then make it possible to calculate an overall $\Delta\Delta G_{\text{bind}}$ between the two ligands by considering each pose of each ligand explicitly. However, because the ligands are not restrained to an orientation during the single topology and dual topologies transformations, reorientations of the ligands may happen “by accident” and need to be filtered out in post-processing if one wants to compute relative free energies of specific ligand orientations. We find the separated topologies method to be the cleanest method of calculating relative free energies between ligands because the sampling is defined explicitly in advance with little need for post-processing and the ligands need not be related either topologically or by orientation.

A notable difference between the three methods is how they handle symmetry of the ligands. Benzimidazole has an axis of symmetry and can bind in two symmetric orientations (an effective symmetry number of 2), whereas each 4-

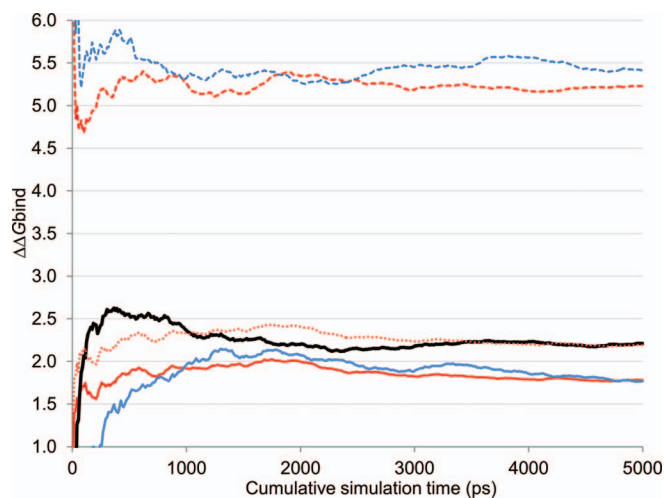


FIG. 4. Free energy estimates for $\Delta\Delta G_{\text{bind}}$ using up to a given amount of simulation (per lambda intermediate) to calculate $\Delta\Delta G_{\text{site}}$. Red: single topology calculations. Blue: dual topologies calculations. Black: our separated dual topologies method. Simulations started from Pose 1 are shown using solid lines; simulations started from Pose 2 are shown in dashes. Our method combines data from simulations started from both poses. Single and dual topology simulations starting from Pose 2 are incorrect by over 3 kcal/mol because the favorable orientation of 4-azaindole has not been sampled. The single and dual topology simulations started from Pose 1 are also incorrect by 0.4 kcal/mol because the symmetric orientation of benzimidazole has not been sampled. In small red dots, we show the single topology calculation started from Pose 1 with a 0.4 kcal/mol symmetry correction added, which leads to a close agreement with our separated topologies (black) method.

azaindole orientation is unique (an effective symmetry number of 1). In the separated topologies method, ligands are restrained during all steps except the restraining steps (steps 1 and 7 in Fig. 2), which prevents sampling of symmetric orientations. As in absolute binding calculations,²⁷ it is unnecessary to sample symmetric orientations during restraining steps, because a symmetry correction can be easily applied. In the case of the transformation from benzimidazole to 4-azaindole, the symmetry correction is effectively a 0.41 kcal/mol increase in $\Delta G_{\text{rest},A}$, which accounts for the fact that the symmetric orientation of benzimidazole was not sampled during this step. This leads to a 0.41 kcal/mol increase in $\Delta\Delta G_{\text{bind}}$.

In the single topology method, benzimidazole must sample both symmetric orientations for the calculation to fully include the change in symmetry between benzimidazole and 4-azaindole. If an indicator function is defined and simulation data decomposed into data collected exclusively in Pose 1 ($\Delta\Delta G_{\text{bind},\text{Pose1}}$) and data collected exclusively in Pose 2 ($\Delta\Delta G_{\text{bind},\text{Pose2}}$), one could add a 0.41 kcal/mol symmetry correction in both $\Delta\Delta G_{\text{bind},\text{Pose1}}$ and $\Delta\Delta G_{\text{bind},\text{Pose2}}$. Because $\Delta\Delta G_{\text{bind},\text{Pose1}}$ and $\Delta\Delta G_{\text{bind},\text{Pose2}}$ differ only the free energy of 4-azaindole, the value of $\Delta\Delta G_{\text{bind}}$ with the more favorable free energy of 4-azaindole— $\Delta\Delta G_{\text{bind},\text{Pose1}}$ —would be taken as the overall $\Delta\Delta G_{\text{bind}}$. In our simulations, single topology simulations started from Pose 1 and Pose 2 did always remain in those poses by inspection (Pose 1 stays in Pose 1; Pose 2 stays in Pose 2). In Fig. 4, we show $\Delta\Delta G_{\text{bind},\text{Pose1}}$ with the added 0.41 kcal/mol symmetry correction in light red dots. This symmetry-corrected single topology result is very close

to the separated topologies result and more correct than the single topology result prior to symmetry correction. A similar decomposition of substates approach could also be taken to correct the dual topologies data. Still, we find the separated topologies method to be the simplest approach to calculating relative free energies in cases with many orientations, because the orientations are defined explicitly in advance and because the ligands are restrained to only sample their specified orientations when decoupled, reducing the need for post-processing.

While we used more total simulation time for the separated topologies method than for the single topology method, convergence seemed to be very rapid (Figure 4). The restraining steps (step 1 and step 7, respectively, Fig. S1³⁰ $\Delta G_{\text{rest,A}}$ and $\Delta G_{\text{rest,B}}$) easily converged to within 0.2 kcal/mol within 1 ns, despite our (unnecessary, but simplifying) use of the inefficient Zwanzig equation (EXP)³¹ to calculate the restraining free energy (Fig. S1).³⁰ The worst converging step for all the Method 3 calculations was the $\Delta G_{\text{charge}} + \Delta G_{\text{vdw,A}}$ step for Pose 2 (Fig. S1).³⁰ We examined these simulations and discovered that late in the trajectories, a water molecule forms a new bridging interaction between the ligand and the protein, which caused the free energy estimate from these specific simulations to converge poorly. Slow events like this can affect all of the methods for computing relative free energies and this does not indicate a convergence challenge unique to the separated topologies method. While the separated topologies method provides a convenient means for computing relative free energies between ligands which do not share an orientation, changes in protein conformation between the bound state of ligand A and the bound state of ligand B may still impede convergence.

CONCLUSION

We present the separated topologies method (Figure 2) for performing relative binding free energy calculations. This method uses two separate ligand topologies, which are restrained to the binding site in individual orientations. Using the example of benzimidazole and 4-azaindole binding to CCP-GA, we demonstrated that the method produces results that are equivalent to two commonly used approaches, when no ligand reorientation is needed. We believe that our approach will be advantageous in cases where the initial and final ligands do not share a similar orientation. Our method also allows for sampling of multiple specific orientations of each ligand and provides a way to combine the results to compute a thermodynamically rigorous overall $\Delta\Delta G_{\text{bind}}$.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support, including: National Science Foundation (NSF) and Department of Defense pre-doctoral fellowships (to G.J.R.); National Institutes of Health (NIH) grant GM096257, NSF EPSCoR Cooperative Agreement No. EPS-1003897, Louisiana Board of Regents Research Competitiveness and Research Enhancement Subprograms and additional support from the Louisiana

Board of Regents (to D.L.M.); and NIH grant GM063592 and GM090205 (to K.A.D.).

APPENDIX: METHODS

Simulation parameters

All production simulation parameters were taken from Ref. 23, which itself is a slightly modified version of the protocol in Ref. 27. On top of this protocol, we also introduced Hamiltonian exchange attempts every 0.04 ps, alternating between exchanging across even and odd lambda transitions. Exchanges were accepted or rejected according to the Metropolis criterion.

System equilibration

Protein Data Bank (PDB) structure 1KXN²⁸ was placed in a truncated dodecahedral box with 1.2 nm between the protein and the box edge. Solvent was added using the GRO-MACS tool genbox and the solvent was equilibrated for 2 ns at constant volume with all protein coordinates held fixed. The protein was then energy minimized and equilibrated for an additional 2 ns, using the procedure in Ref. 27. The benzimidazole ligand was then inserted into the equilibrated, solvated protein using the coordinates in PDB structure 1KXM²⁸ and all water molecules within 2.0 Å of the ligand were removed. This simulation was then equilibrated for an additional 1 ns. The initial 4-azaindole coordinates were taken from the equilibrated benzimidazole structure, using the Pose 1 and Pose 2 atom to atom mappings.

Free energy calculations

Method 1.

Benzimidazole was transformed into 4-azaindole in the protein using a Hamiltonian exchange free energy calculation with 8 total lambda intermediates. Within these eight intermediates, the charge, Lennard-Jones, and bonded parameters were perturbed simultaneously. The intermediates used were as follows (charge, VdW, bonded): (0, 0, 0), (0.25, 0, 0), (0.5, 0, 0), (0.75, 0, 0), (1.0, 0.25, 0.25), (1.0, 0.5, 0.5), (1.0, 0.75, 0.75), (1.0, 1.0, 1.0). An identical procedure was used for the transformation in solution, starting from a pre-equilibrated structure of benzimidazole in water. Free energy differences were analyzed using BAR.²⁵

Method 2.

The two ligands were attached to each other at the atoms indicated in Figure 3 using a 10 kcal/mol/Å² restraint and exclusions were used to eliminate all non-bonded interactions between the two ligands. Two sets of Hamiltonian exchange free energy calculations were used to convert benzimidazole into 4-azaindole in the protein. Starting from a simulation with benzimidazole fully coupled to the protein, the Lennard-Jones interactions of 4-azaindole were coupled to the protein using a Hamiltonian exchange^{20,21} free energy calculation with 4 lambda intermediates, spaced at $\lambda = 0.0, 0.33,$

0.67, and 1.00. In a second, separate Hamiltonian exchange free energy calculation with 8 intermediates, the charges on 4-azaindole were created simultaneously with the charges on benzimidazole being annihilated, followed by the Lennard-Jones parameters on benzimidazole being decoupled from the protein. This 8-intermediate set of simulations used an identical lambda schedule as in Method 1, although no bonded parameters were perturbed. Free energy differences were analyzed using BAR and Hamiltonian exchange was applied as in Method 1. To calculate $\Delta\Delta G_{\text{solv}}$, the hydration free energy of each compound was calculated using the method in Ref. 32. No corrections for net charge changes were applied, which would approximately cancel for a relative hydration free energy.

Method 3.

Each ligand was restrained to the binding site using C β , C α , and N of His175 on the protein as a, b, and c, respectively, in the scheme of Boresch *et al.*²² These atoms were chosen because they near the ligand, are rigidly and firmly oriented relative to the binding site, and because they allow the angles $\angle\text{baA}$ $\angle\text{aAB}$ in Fig. 2 of Ref. 22 to be near 90° (angles far from 90° create wildly fluctuating dihedral angles). Restraints were set at 10 kcal/mol/rad² and 10 kcal/mol/Å². Benzimidazole was transformed into 4-azaindole using the same two sets of Hamiltonian exchange simulations and lambda schedules as in Method 2. A separate, unrestrained simulation of each ligand in the binding site was used to calculate the restraining energy, which was analyzed using the Zwanzig equation. $\Delta\Delta G_{\text{solv}}$ was calculated as in Method 2.

Error analysis

Data for BAR calculations was subsampled and intervals of the statistical inefficiency of the work (W) time series and statistical uncertainties shown in Table I were calculated using the method of Shirts and Chodera.²⁶ Statistical uncertainties in Zwanzig equation calculations shown in Table I (the restraining energies for Method 3) were calculated from the standard error of the mean of $\exp(-\beta^{-1} W)$ and data was not subsampled.

Software

Free energy calculations in the protein-ligand complex and the Method 1 simulations of the ligand in solution were performed using GROMACS 4.5.3-dev-20110305-9e524, kindly provided by Shirts. The hydration free energy calculations for the ligands in solution for Method 2 and Method 3 were performed using GROMACS 4.0.7.³³

Force fields

The protein was modeled from PDB structure 1KXN²⁸ using the AMBER99SB force field,³⁴ solvated in TIP3P water. The generalized amber force field (GAFF)³⁵ generated using Antechamber 1.27 and AM1-BCC partial charges^{36,37}

generated using the OEChem toolkit version 1.5 were used for 4-azaindole and benzimidazole. Heme parameters and partial charges were taken from the Hemoglobin model included with Amber 8, except with an additional +1 charge added to the iron to create Fe(III), as in Banci *et al.*³⁸ Additional parameters were added to create the Fe–N ϵ bond between the Heme and His175, and the four Heme N–Fe–N ϵ angles, using force constants taken from Banci *et al.*³⁸ The equilibrium bond length and angles were generated based on the average of existing crystal structures of CCP W191G. To improve convergence and to avoid conflating convergence problems caused by protein conformational changes with convergence problems caused by ligand reorientation, we restrained the following protein dihedral angles (according to the numbering in PDB 1KXN) in all simulations, using force constants of 10 kcal/mol/Å²: Gly190 N–C α –C–Gly191 N (167°), Leu202 C–Asn203 C' α –C β –C γ (–60°), Heme C2A–C3A–CAA–CBA (89°), Heme C3D–CAD–CBD–CGD (167°), and Heme C2D–C3D–CAD–CBD (76°). In simulations without these restraints, these dihedral angles rotate once every several nanoseconds, which can affect the free energy difference between the ligands. The restraints were applied to ensure that sampling of these angles was identical between all three methods. Protonation states of titratable residues were selected by running MCCE (Multi-Conformation Continuum Electrostatics) 2.2,^{39,40} on PDB structure 1KXN²⁸ with the Heme removed at pH 4.5.

- ¹D. L. Mobley and P. V. Klimovich, *J. Chem. Phys.* **137**, 230901 (2012).
- ²J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande, *Curr. Opin. Struct. Biol.* **21**, 150 (2011).
- ³J. Michel and J. W. Essex, *J. Comput.-Aided Mol. Des.* **24**, 639 (2010).
- ⁴C. F. Wong and J. A. McCammon, *J. Am. Chem. Soc.* **108**, 3830 (1986).
- ⁵S. Miyamoto and P. A. Kollman, *Proteins* **16**, 226 (1993).
- ⁶J. W. Essex, D. L. Severance, J. Tirado-Rives, and W. L. Jorgensen, *J. Phys. Chem. B* **101**, 9663 (1997).
- ⁷T. Fox, T. S. Scanlan, and P. A. Kollman, *J. Am. Chem. Soc.* **119**, 11571 (1997).
- ⁸M. L. Plouffe Price and W. L. Jorgensen, *J. Am. Chem. Soc.* **122**, 9455 (2000).
- ⁹B. C. Oostenbrink, J. W. Pitera, M. M. H. van Lipzig, J. H. N. Meerman, and W. F. van Gunsteren, *J. Med. Chem.* **43**, 4594 (2000).
- ¹⁰A. Aleksandrov and T. Simonson, *Biochemistry* **47**, 13594 (2008).
- ¹¹D. Jiao, J. Zhang, R. E. Duke, G. Li, M. J. Schnieders, and P. Ren, *J. Comput. Chem.* **30**, 1701 (2009).
- ¹²S. E. Boyce, D. L. Mobley, G. J. Rocklin, A. P. Graves, K. A. Dill, and B. K. Shoichet, *J. Mol. Biol.* **394**, 747 (2009).
- ¹³L. Wang, B. J. Berne, and R. A. Friesner, *PNAS* **109**, 1937 (2012).
- ¹⁴J. Michel, M. L. Verdonk, and J. W. Essex, *J. Chem. Theory Comput.* **3**, 1645 (2007).
- ¹⁵J. Michel and J. W. Essex, *J. Med. Chem.* **51**, 6654 (2008).
- ¹⁶S. Riniker, C. D. Christ, N. Hansen, A. E. Mark, P. C. Nair, and W. F. van Gunsteren, *J. Chem. Phys.* **135**, 024105 (2011).
- ¹⁷S. Banba and C. L. Brooks, *J. Chem. Phys.* **113**, 3423 (2000).
- ¹⁸S. Banba, Z. Guo, and C. L. Brooks, *J. Phys. Chem. B* **104**, 6903 (2000).
- ¹⁹D. L. Mobley and P. Klimovich, *J. Chem. Phys.* **137**, 230901 (2012).
- ²⁰C. J. Woods, J. W. Essex, and M. A. King, *J. Phys. Chem. B* **107**, 13703 (2003).
- ²¹Y. Sugita, A. Kitao, and Y. Okamoto, *J. Chem. Phys.* **113**, 6042 (2000).
- ²²S. Boresch, F. Tettlinger, M. Leitgeb, and M. Karplus, *J. Phys. Chem. B* **107**, 9535 (2003).
- ²³D. L. Mobley, A. P. Graves, J. D. Chodera, A. C. McReynolds, B. K. Shoichet, and K. A. Dill, *J. Mol. Biol.* **371**, 1118 (2007).
- ²⁴J. Wang, Y. Deng, and B. Roux, *Biophys. J.* **91**, 2798 (2006).
- ²⁵C. H. Bennett, *J. Comput. Phys.* **22**, 245 (1976).

- ²⁶M. R. Shirts and J. D. Chodera, *J. Chem. Phys.* **129**, 124105 (2008).
- ²⁷D. L. Mobley, J. D. Chodera, and K. A. Dill, *J. Chem. Phys.* **125**, 084902 (2006).
- ²⁸R. Rosenfeld, A. Hays, R. Musah, and D. Goodin, *Protein Sci.* **11**, 1251 (2002).
- ²⁹P. Nuno Palma, M. João Bonifácio, A. Isabel Loureiro, and P. Soares-da-Silva, *J. Comput. Chem.* **33**, 970–986 (2012).
- ³⁰See supplementary material at <http://dx.doi.org/10.1063/1.4792251> for convergence plots for all alchemical steps.
- ³¹M. R. Shirts and V. S. Pande, *J. Chem. Phys.* **122**, 144107 (2005).
- ³²D. L. Mobley, É. Dumont, J. D. Chodera, and K. A. Dill, *J. Phys. Chem. B* **111**, 2242 (2007).
- ³³B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *J. Chem. Theory Comput.* **4**, 435 (2008).
- ³⁴V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Proteins* **65**, 712 (2006).
- ³⁵J. Wang, W. Wang, P. A. Kollman, and D. A. Case, *J. Mol. Graphics Modell.* **25**, 247 (2006).
- ³⁶A. Jakalian, D. B. Jack, and C. I. Bayly, *J. Comput. Chem.* **23**, 1623 (2002).
- ³⁷A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, *J. Comput. Chem.* **21**, 132 (2000).
- ³⁸L. Banci, P. Carloni, and G. G. Savellini, *Biochemistry* **33**, 12356 (1994).
- ³⁹E. G. Alexov and M. R. Gunner, *Biophys. J.* **72**, 2075 (1997).
- ⁴⁰R. E. Georgescu, E. G. Alexov, and M. R. Gunner, *Biophys. J.* **83**, 1731 (2002).