THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

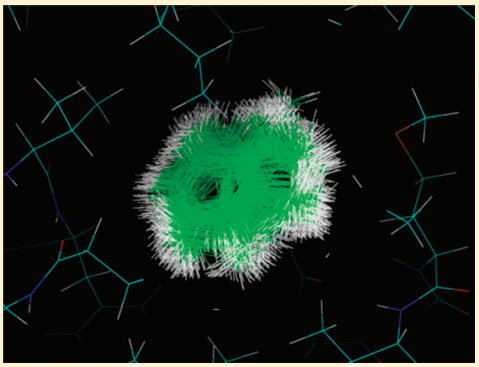
pubs.acs.org/JPCB

Protein−Ligand Binding Free Energies from Exhaustive Docking

Enrico O. Purisima* and Hervé Hogues

Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada

ABSTRACT: We explore the use of exhaustive docking as an alternative to Monte Carlo and molecular dynamics sampling for the direct integration of the partition function for protein−ligand binding. We enumerate feasible poses for the ligand and calculate the Boltzmann factor contribution of each pose to the partition function. From the partition function, the free energy, enthalpy, and entropy can be derived. All our calculations are done with a continuum solvation model that includes solving the Poisson equation. In contrast to Monte Carlo and molecular dynamics simulations, exhaustive docking avoids (within the limitations of a discrete sampling) the question of "Have we run long enough?" due to its deterministic complete enumeration of states. We tested the method on the T4 lysozyme L99A mutant, which has a nonpolar cavity that can accommodate a number of small molecules. We tested two electrostatic models. Model 1 used a solute dielectric of 2.25 for the complex apoprotein and free ligand and 78.5 for the solvent. Model 2 used a solute dielectric of 2.25 for the complex and apoprotein but 1.0 for the free ligand. For our test set of eight molecules, we obtain a reasonable correlation with a Pearson $r^2 = 0.66$ using model 1. The trend in binding affinity ranking is generally preserved with a Kendall $\tau = 0.64$ and Spearman $\rho = 0.83$. With model 2, the correlation is improved with a Pearson $r^2 = 0.83$, Kendall $\tau = 0.93$, and Spearman $\rho = 0.98$. This suggests that the energy function and sampling method adequately captured most of the thermodynamics of binding of the nonpolar ligands to T4 lysozyme L99A.



INTRODUCTION

The ability to accurately predict the strength of intermolecular associations is of great scientific and practical importance. It would render the practice of structure-based drug design more robust and accelerate the pace of drug discovery. A number of approaches have been developed over the years that range from the more empirical and knowledge-based to the more force field-based methods.[1] Among the latter, the end-point methods[2] are attractive in terms of their simplicity and computational efficiency relative to more rigorous perturbation or alchemical simulations.[3] In our own lab we have previously developed one such end-point method, the solvated interaction energy (SIE) model,[4,5] which has been moderately successful in predicting protein−ligand binding affinities.[6,7] However, SIE largely ignores entropy and attempts to capture its contribution by empirically scaling what is largely an enthalpy term, i.e., it relies on entropy−enthalpy compensation. This works in an average gross way but in general is too crude for precise binding affinity calculations. In this paper we present a method that explicitly incorporates entropic effects while remaining entirely within the realm of an end-point method.

The challenge of force field- or physics-based methods is to properly model the thermodynamics of binding in order to obtain the free energies correctly. The partition function of a system is a fundamental quantity from which all of its thermodynamics can be derived.[8] However, the direct calculation of the partition function has traditionally been difficult due to problems of convergence. Alchemical and free energy perturbation methods have been developed to bypass the problem by not calculating the partition function directly.[3] Nevertheless, there are methods that have been developed to integrate the partition function directly that are based largely on thorough stochastic sampling of phase space.[9−11] Here, we explore the use of exhaustive docking as an alternative to Monte Carlo and molecular dynamics sampling for the direct integration of the partition function. We enumerate feasible poses for the ligand and calculate the Boltzmann factor contribution of each pose to the partition function. From the partition function, the free energy, enthalpy, and entropy can be derived. All our calculations are done with a continuum solvation model based on solving the Poisson equation.[12,13] In contrast to Monte Carlo and molecular dynamics simulations, exhaustive docking avoids (within the limitations of a discrete sampling) the question of "Have we run long enough?" due to its deterministic complete enumeration of states. Salaniwal et al.[14] have previously described a partition function-based scoring method for virtual screening that makes use of a grid-based exhaustive search. However, their formulation was designed more for speed rather than accuracy, and they introduced some rather drastic approximations unsuitable for detailed binding free energy calculations. For example, they

Special Issue: Harold A. Scheraga Festschrift

Received: December 31, 2011
Revised: March 16, 2012
Published: March 20, 2012

dx.doi.org/10.1021/jp212646s | J. Phys. Chem. B 2012, 116, 6872−6879

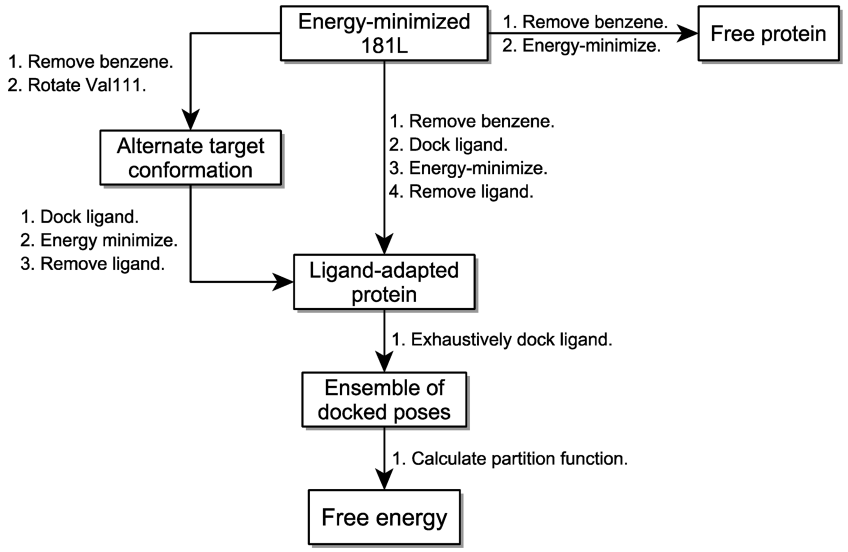


Figure 1. Flowchart of steps in the calculation.

assumed that low-energy conformers below a certain cutoff were all equi-energetic.[14] Here we build the partition function using detailed energies calculated for each pose using the AMBER[15] and GAFF[16] force fields supplemented by a high-quality continuum solvation model.[5,12,13]

To test the method we wanted a simple but not trivial system. The system had to satisfy the following: (1) measured binding affinities for a series of relatively rigid ligands, (2) known binding mode for each of the ligands, (3) high-resolution structure of the target with no large conformational changes from one ligand to another. The T4 lysozyme L99A mutant satisfies these criteria.[17] The mutation creates a nonpolar cavity that can accommodate a number of small molecules. The cavity has the added advantage that it is entirely buried and unsolvated. This is unlike a typical enzyme active site where ligand binding requires extensive water molecule displacement. This simplification makes T4-Lysozyme L99A an attractive model system for testing binding free energy prediction algorithms. It has been used in a number of published studies on free energy calculations.[18−24]

In this work, we exhaustively docked a set of eight small molecules (benzene, benzofuran, indene, isobutylbenzene, indole, *n*-butylbenzene, *o*-xylene and *p*-xylene) in the T4 lysozyme L99A binding cavity and used the poses to build up the partition function for the complex. The resulting binding free energies were in reasonable agreement with experimental values. Moreover, decomposition into enthalpic and entropic components could be carried out. We observed a pattern of entropy−enthalpy compensation, i.e., anticorrelation, that has been observed in other systems.[25]

METHODS

Structure Preparation. The crystal structure (PDB code 181L) of T4 lysozyme L99A in complex with benzene was taken as the starting point for our simulations. All water molecules were removed, hydrogen atoms were added, and AMBER and GAFF force field parameters were assigned to the protein and ligand, respectively. The complex was energy minimized with a distance-dependent dielectric function ($\varepsilon = 4r$) to an rms gradient of 0.01 kcal/mol. A set of residues around the binding cavity was then defined as the "hotset" that would be allowed to move during subsequent energy minimizations while the rest of the protein was held fixed. Figure 1 shows schematically the subsequent steps in structure preparation. The benzene ligand was removed from the energy-minimized 181 L protein−ligand complex. A rectangular box enclosing the entire binding cavity was constructed. Each ligand (benzene, benzofuran, indene, and indole) was then docked into the binding cavity using the program WILMA (see below) and complex energy-minimized. As a result, each ligand has a modeled cognate target protein structure that can differ structurally and energetically from the others in the hotset region. This is meant to capture to some degree the energetics of induced fit effects in the protein. The modeled cognate protein was then the starting point for exhaustive docking to generate poses for building up the partition function for the bound state. For isobutylbenzene, *n*-benzene, *p*-xylene, and *o*-xylene it is known that the cognate protein structure has the Val111 side chain in a different rotameric state. Hence, the modeled cognate target proteins for these ligands were generated by first rotating the Val111 side chain prior to WILMA docking and energy-minimization. To generate the free state of the protein, the original energy-minimized 181 L was energy-minimized further in the absence of any ligand allowing only the hotset atoms to move. By keeping large parts of the protein fixed across all complexes and the free state, we reduce the introduction of noise due to changes in parts of the protein far away from the binding site. Such changes can be quite unpredictable and are due to the chaotic nature of energy minimization.[26]

Ligand docking. Docking was carried out using an in-house exhaustive docking software called WILMA (manuscript in preparation) that is similar in spirit to the docking program FRED[27,28] (OpenEye Scientific Software, Santa Fe, New Mexico). WILMA uses a brute-force searching approach that scores all the discrete rotational and translational states of a rigid ligand within a defined binding site region of a rigid protein. Using an efficient filtering method, the program exhaustively enumerates, scores, and ranks all the ligand poses that do not overlap with the protein. It uses a weighted five-term scoring function that was trained to recover the native binding modes of crystal structures of protein−ligand

Table 1. Experimental and Calculated Binding Free Energies[a]

		electrostatic model 1[a]			electrostatic model 2[b]		
	expt	ΔG	$\Delta\langle E\rangle$	$-T\Delta S$	ΔG	$\Delta\langle E\rangle$	$-T\Delta S$
benzene	−5.2	−6.45 (0.01)	−16.78 (0.06)	10.33 (0.06)	−5.32 (0.01)	−15.61 (0.06)	10.29 (0.06)
benzofuran	−5.5	−8.72 (0.07)	−21.02 (0.27)	12.31 (0.21)	−6.99 (0.06)	−19.28 (0.27)	12.28 (0.21)
indene	−5.1	−6.47 (0.07)	−18.58 (0.18)	12.11 (0.13)	−5.14 (0.06)	−17.24 (0.18)	12.09 (0.13)
i-butylbenzene	−6.5	−8.31 (0.25)	−22.08 (0.37)	13.77 (0.18)	−7.35 (0.20)	−21.12 (0.37)	13.77 (0.18)
indole	−4.9	−6.89 (0.90)	−19.04 (1.26)	12.15 (0.62)	−3.47 (0.64)	−15.56 (1.26)	12.09 (0.63)
n-butylbenzene	−6.7	−8.40 (0.05)	−22.10 (0.09)	13.69 (0.05)	−7.50 (0.04)	−21.20 (0.09)	13.70 (0.05)
p-xylene	−4.7	−4.63 (0.34)	−16.14 (0.37)	11.52 (0.15)	−3.63 (0.25)	−15.13 (0.37)	11.51 (0.15)
o-xylene	−4.6	−4.41 (0.03)	−16.12 (0.28)	11.70 (0.13)	−3.38 (0.19)	−15.08 (0.28)	11.69 (0.12)

[a]All values are in kcal/mol. The calculated average energies and entropic contributions are shown as well. Standard deviations over triplicate runs are shown in parentheses. [a]D_{in} = 2.25 for the complex, apoprotein and free ligand. D_{out} = 78.5. [b]D_{in} = 2.25 for the complex and apoprotein. D_{in} = 1.0 for the free ligand.

complexes. The scoring function includes a 6−12 Lennard-Jones potential, a Coulomb interaction term, an explicit H-bond term, which considers donor and acceptor orientations, and two surface and polar-surface complementarity terms. Docking is done within a predefined rectangular volume (10 Å × 11 Å × 12 Å) with a translation step size of 0.25 Å. This step size was a compromise between having a high-resolution grid and having a three-dimensional (3D) grid matrix with reasonable memory requirement. In docking studies, we have found this grid spacing to be sufficient to reproduce binding modes found in crystal structures in the Protein Data Bank.

Rigid-body rotations were were carried out systematically as follows. First a rotation axis was selected from a unit sphere with points spread uniformly across the surface. The long axis of the ligand was aligned with the chosen rotation axis and followed by a rotation about that axis. The discrete rotation increment of the ligand was adjusted to ensure that the maximum movement of any atom between adjacent orientations was less than 0.25 Å. By systematically sampling various rotation axes and stepping through rotation increments, the rotational states of the ligand were exhaustively enumerated. WILMA was developed for use in a virtual screening pipeline and hence is very quick in generating a ranked ensemble of poses. It was recently used successfully in SAMPL-3, a blind test for assessing algorithms for virtual screening and affinity prediction.[6]

Torsional Sampling. Six of the eight ligands are essentially rigid, while isobutylbenzene and *n*-butylbenzene have two and three rotatable bonds, respectively. We do not count rotations of methyl groups. The local conformational space of these ligands was sampled around the bioactive conformation by systematically scanning the dihedral angles of the free ligand within a ±30° range in 15° increments and keeping only sterically allowed conformations. This led to 17 and 47 conformations for isobutylbenzene and *n*-butylbenzene, respectively. All these conformations were docked exhaustively in the binding cavity to generate the ensembles of bound poses.

Energy Function. The energy of the system has two components. The first is the molecular mechanics energy of the complex calculated using the AMBER force field[15] for the protein ad GAFF[16] for the ligand. The second term is the solvation free energy consisting of the electrostatic reaction field energy and a surface-area-based nonpolar solvation term. We used solute and solvent dielectric constants of 2.25 and 78.5, respectively. These are the default values that seem to work well with the SIE scoring function.[4,5,7,29] The reaction field energies were calculated using the BRI BEM program[12,13] and a variable probe molecular surface[30,31] (VP-MS) to define the dielectric boundary. The Born radii were taken to be 1.1 times the AMBER van der Waals radii. The nonpolar solvation free energy (in kcal/mol) is calculated as 0.0129 × Area + 0.55. We also tried a combination of solute dielectric constants in which the complex and the protein were assigned a dielectric constant of 2.25, while the free ligand was assigned a dielectric constant of 1.0.

Partition Function Calculation. The partition function for the protein−ligand complex was built up directly from a discrete summation of Boltzmann factors over the exhaustive search poses multiplied by the translational and rotational grid dimensions. This is a discrete numerical approximation of the integral of the Boltzmann factor over phase space.

$$Z^{PL} = 8\pi^2 V^0 \sum_{ijk} e^{-\beta(E_{ijk}^{PL}+W_{ijk}^{PL})}(\Delta r)^3(8\pi^2/n_{orient}) \quad (1)$$

There are two energy terms in the exponential. These are contributions from the molecular mechanics energy, E, and the solvation free energy, W. The summation is over the poses corresponding to the j and k translational and rotational states of the bound ligand and over i for rotameric states of the ligand if they exist. β is $(k_BT)^{-1}$ where k_B is Boltzmann's constant and T = 300 K. Δr is 0.25 Å and is the length of the edge of a cubic volume of translational phase space associated with a grid point. n_{orient} is the number of rotational states sampled, i.e., it is the product of the number points on a unit sphere defining the sampled axis of rotation multiplied by the number of angular increments about the rotation axis. Each rotational state represents a volume of $8\pi^2/n_{orient}$ in rotational phase space. $8\pi^2$ comes from integrating over 4π radians (the space of orientations of the rotation axis) multiplied by 2π radians (the range for rotation about the axis). Note that the various poses sample rotational phase space in equivolume chunks, and the $8\pi^2/n_{orient}$ factor has any Jacobian factors implicitly built-in. (J dΩ is discretized into J $\Delta\Omega = 8\pi^2/n_{orient}$, where J is the Jacobian and J dΩ is an infinitesimal volume of phase space.) The entire protein−ligand complex itself has translational and rotational rigid-body degrees of freedom, which, if integrated out for a 1 M solution, gives the $8\pi^2V^0$ factor preceding the summation where V^0 = 1660 Å.

dx.doi.org/10.1021/jp212646s | J. Phys. Chem. B 2012, 116, 6872−6879

We can construct analogous partition functions for the free ligand and free protein and obtain an expression for the free energy of binding.

$$\Delta G_{\text{bind}} = -RT \ln\left[\frac{\sum_{ijk} e^{-\beta(E_{ijk}^{PL}+W_{ijk}^{PL})}(\Delta r)^3(8\pi^2/n_{\text{orient}})}{8\pi^2 V^0[\sum_i e^{-\beta(E_i^L+W_i^L)}][e^{-\beta(E^P+W^P)}]}\right] \quad (2)$$

The superscripts PL, L, and P refer to the protein−ligand complex, free ligand, and free protein moieties, respectively. The free protein is approximated as a single state. The partition functions of the free ligand and free protein each have a factor of $8\pi^2V^0$ corresponding to unhindered translational and rotational degrees of freedom. Together with eq 1, this leads to a single factor of $8\pi^2V^0$ in the denominator of eq 2. A thorough discussion of the statistical thermodynamics of computing binding affinities can be found in these reviews.[32,33] In our exhaustive docking, we consider symmetry-related poses as distinct poses. Hence, no symmetry correction terms are included in the partition function expressions.

RESULTS AND DISCUSSION

We see from eq 2 that the main task is to enumerate the bound poses, calculate their Boltzmann factors, and sum them up. In carrying out the summation for the complex, we included up to the first 2000 top-ranked poses returned by WILMA. As we shall see later, WILMA ranks the poses quite well such that the predominant poses that contribute most to the partition function are near the top of the list. Hence, for this system, using 2000 poses appears to be more than sufficient. Table 1 summarizes the results averaged from triplicate runs for the binding free energies obtained from exhaustive docking. Each replicate was carried out using rectangular grid boxes that were slightly displaced relative to each other. Two sets of results are presented corresponding to two electrostatic models (essentially the choice of solute dielectric constant). Electrostatic model 1 uses a solute dielectric of 2.25 for the protein, ligand, and complex and 78.5 for the solvent. These is the default combination that we have found useful in SIE calculations.[4,5,7,29] Electrostatic model 2 uses a solute dielectric of 2.25 for the protein and complex but a dielectric of 1.0 for the free ligand. The solvent dielectric was 78.5 as well. With model 1, the correlation of binding free energies with experiment has a Pearson r^2 = 0.66. The ranking is relatively well preserved with a Kendall τ = 0.64 and Spearman ρ = 0.83. With model 2, the Pearson correlation improves to r^2 = 0.83. The experimental ranking is also better reproduced with a Kendall τ = 0.93 and Spearman ρ = 0.98. The agreement with experiment is surprisingly good given the approximations taken in the method. The mean unsigned error (MUE) is 1.45 kcal/mol for model 1 and 0.88 kcal/mol for model 2. Figure 2 is a scatter plot of the predicted versus experimental binding free energies. The dynamic range of the calculated values is about 1.5−2 times larger than the experimental one, which is rather narrow at 2.1 kcal/mol. However, the ligands at the extremes of the range are correctly identified from the calculated free energies and the overall trend in terms of rank is reproduced reasonably. This suggests that the energy function and sampling method adequately captured most of the thermodynamics of binding of the nonpolar ligands to T4-lysozyme L99A. Note that our method neglects changes in protein configurational entropy. We expect that binding of the ligands induces some

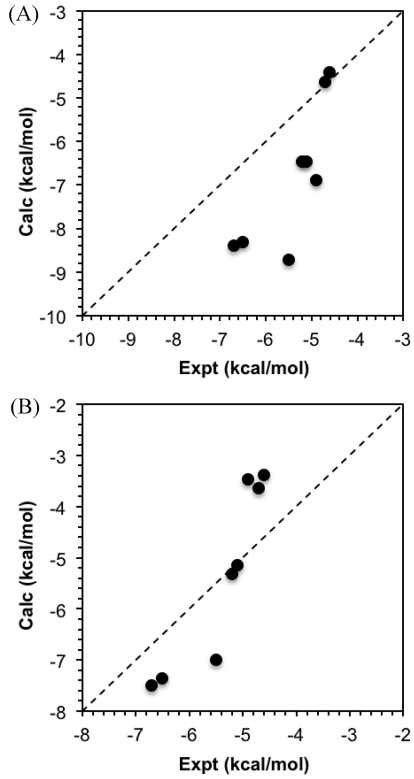


Figure 2. Scatter plots of predicted binding free energies versus experimental values. The $y = x$ diagonal is shown for reference. (A) Electrostatic model 1. (B) Electrostatic model 2.

rigidification around the binding cavity resulting in some additional entropic cost. This may be more pronounced for the larger ligands such as *n*-butylbenzene and *i*-butylbenzene.

The special properties of the binding cavity and the rigid nature of the ligands likely helped in making this system tractable for our method. The binding cavity is completely buried and desolvated to begin with. Because we keep the protein rigid in our docking simulation, different poses of a nonpolar ligand in the cavity hardly alter the solvation free energy of the complex. The partial charges of the buried ligand modify the reaction field energy just slightly. Hence, the only solvation effect we have to worry about is the desolvation of the ligands. Our previous experience with calculating hydration free energies of small molecules[34,35] suggests that a continuum solvation model can correctly estimate the desolvation free energy of the rigid nonpolar ligands in this study. It has previously been pointed out that continuum solvation models that use a molecular surface rather than a solvent-accessible surface for T4 lysozyme L99A mistakenly treat the binding cavity as solvated.[36] However, in our case, we use a VP-MS,[30] which uses different probe radii for anionic, polar, and nonpolar surfaces. A larger probe radius of 1.6 A is used for surfaces associated with nonpolar atoms. Because nonpolar atoms line the binding cavity, the larger probe radius causes the binding cavity to be fully desolvated even without any ligand in it, correctly mimicking the desolvated native state of the cavity.

The rigid nature of the ligands also simplifies our calculations. In ligand binding, one can formally think of the ligand as losing its translational and rotational degrees of freedom, with the complex gaining new vibrational modes. These new vibrational modes are usually modeled using normal-mode analysis methods. However, our method does

dx.doi.org/10.1021/jp212646s | J. Phys. Chem. B 2012, 116, 6872−6879

not deal with vibrations directly. On the other hand, since our ligands are largely rigid ring molecules, only high-frequency modes due to bond stretching and angle bending are present in solution aside from translation and rotation. These high-frequency modes are largely unchanged in the complex. Upon binding, the new "vibrational modes" in the complex are essentially the residual translational and rotational motion of the ligand (within a rigid protein approximation), which is sampled quite well by the exhaustive docker. For the more flexible ligands, isobutylbenzene and *n*-butylbenzene, the torsional modes were sampled by docking a limited set of conformations for which the torsion angles were scanned in a ±30° range in 15° increments. This will capture changes in the steepness of the torsional well between the free and bound states but will not account for the existence of other wells.

Despite the seeming simplicity of the T4 lysozyme system, it is not a trivial exercise to compute binding free energies for its ligands. For example, Mobley et al.[18] point out that multiple orientations exist for the ligands, and these are difficult to sample in a molecular dynamics simulation when starting from a single orientation. Using their alchemical free energy methods, they obtained improved results when they included contributions from multiple orientations. Because ligands can get trapped near their starting orientation for the duration of an MD simulation, several simulations starting from different docked poses had to be carried out. In contrast, the very nature of exhaustive docking overcomes this problem as all relevant poses are enumerated.

Strain Energy. As described in the Methods section, a hotset region around the L99A cavity is allowed to relax during energy minimization with each ligand to produce an initial cognate binding site for use with exhaustive sampling. Hence, depending on the ligand being docked, the internal energy of the target protein is different. This difference in internal energy is included the energy of the complex that is used in the partition function. The same hotset region will also have a slightly altered conformation in the apoprotein with a corresponding change in energy. Table 2 summarizes the difference in internal energies of the target proteins relative to the apoprotein as well as relative to the target protein used for docking benzene (PDB code 181L). The strain energies are composed of the changes in the AMBER energies and the solvation energies. The benzene ligand induces the smallest strain energy, while the bulkiest ligands, *n*-butylbenzene and isobutylbenzene, produce the largest strain energies, part of which is due to the rotation of the Val111 side chain to accommodate them.

Table 2. Contribution of Protein Strain Energies[a]

	amber	solvation	ΔE	rel. ΔE
benzene	1.78	0.22	1.99	0.00
benzofuran	3.48	0.35	3.83	1.84
indene	7.24	0.76	7.99	6.00
i-butylbenzene	9.96	0.18	10.13	8.14
indole	4.84	−0.06	4.77	2.78
n-butylbenzene	10.79	0.57	11.36	9.37
p-xylene	7.27	0.61	7.88	5.89
o-xylene	7.53	0.66	8.19	6.20

[a]All values are in kcal/mol. The amber, solvation, and ΔE values (columns 2−4) are relative to the apoprotein.

Val111 Conformational Change. Binding of four of the ligands (isobutylbenzene, *n*-butylbenzene, *o*-xylene, and *p*-xylene) is known to induce a conformational change in the Val111 side chain in the binding cavity relative to the apoprotein.[17] The rotation is required to avoid steric interference with the ligands. Hence, pose generation for these ligands was carried out with the Val111 side chain prerotated. The Val111 side chain was also rotated for indene in order to relieve steric clashes. As pointed out by Deng et al.,[19] the crystal structure with indene (PDB code 183L) is ambiguous as to the rotameric state of the Val111 side chain. The energetic consequence of these side chain rotations is built into the modeled cognate protein structure for each ligand. We see that the ligands with rotated Val111 side chains have the highest strain energies in their target proteins (Table 2). The energy in the Boltzmann factors for the partition function of the complex includes the internal energy of the protein, which is different for each ligand due to the hotset residues that were allowed to optimize in modeling the cognate protein. We used modeled cognate proteins rather than the actual available crystal structures in order to reduce the noise arising from differences in protein coordinates. The use of a modeled cognate structure was a compromise in taking into account some level of induced fit while allowing cancellation of internal energies for most of the protein between the complex and apo form.

Mobley et al. found that keeping the protein rigid is quite detrimental to estimating binding free energies.[18] They were able to improve agreement with experiment by performing separate local minimizations of the complex for each ligand prior to simulations with a rigid protein.[18] However, these local minimizations were still not sufficient to obtain good agreement with experimental values. The impact of the use of a rigid or partially rigid protein is greater in the context of molecular dynamics sampling because a ligand can get trapped in a particular orientation if concerted motion with the protein is required to pass from one orientation to another. With an exhaustive docker, ligand orientations compatible with a given rigid protein conformation can be sampled directly without the need for a low-barrier pathway.

Efficiency of WILMA Ranking. The poses returned by the WILMA docking program are ranked according to its internal scoring function, with the first pose in the list being the most favorable. We do not expect the docking score to correlate perfectly with the full AMBER + solvation energy for a pose. However, we do expect the top-ranked poses from WILMA to be enriched in low-energy poses even after rescoring with a more accurate energy function. This means the contribution to the partition function (and hence free energy) will be dominated by the early poses in the list, allowing the partition function to be calculated with a shorter list of poses. Figure 3A shows a plot of the Boltzmann factor for each pose for *n*-butylbenzene normalized against the lowest energy pose for one of the replicates. We see that most of the poses that have a relative Boltzmann factor greater than 0.001 are found early on in the list. In fact, we have 90% of the partition function after including just the first 17 poses. We have 99% of the partition function after the first 111 poses. Using just the first 17 and 111 poses for the partition function translates into a free energy error of just 0.06 and 0.006 kcal/mol, respectively. We see similar behavior for the more rigid *p*-xylene ligand (Figure 3B). We obtain 90% of the partition function after the first 13 poses and 99% after the first 40. Hence, we can take advantage of the

dx.doi.org/10.1021/jp212646s | J. Phys. Chem. B 2012, 116, 6872−6879

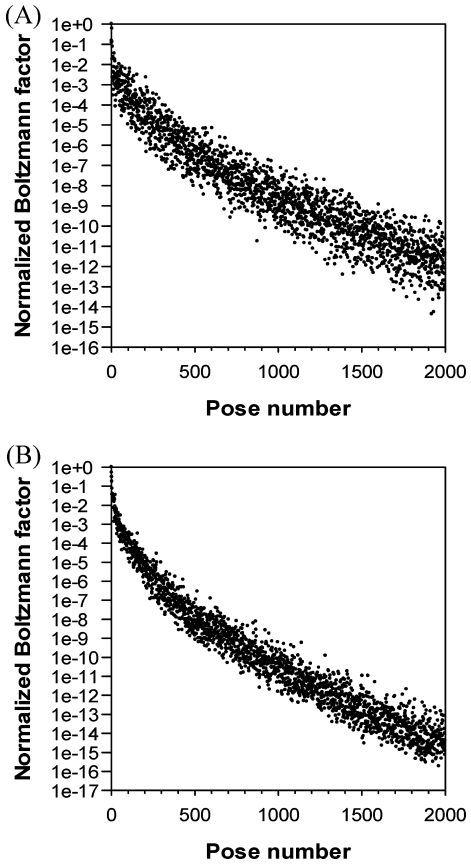


Figure 3. Relative Boltzmann factor versus pose number: (A) *n*-butylbenzene; (B) *p*-xylene. A logarithmic scale is used for the vertical axis. The poses are ordered according to their docking score. Those with the highest docking score (i.e., low pose number) contribute most to the partition function.

ranking provided by WILMA to speed up the partition function calculation with only a minor sacrifice in accuracy. It takes WILMA about a second to provide the 2000 poses for lysozyme, while it takes about 10 s to calculate AMBER + solvation energy for each pose, most of the cost due to solving the Poisson equation for the reaction field energy. Of course, in this particular system the solvation free energy calculation for the complex could have been approximated as a constant because the binding site is completely buried and desolvated, making the solvation free energy of the complex largely insensitive to the ligand pose. However, for a more general case where the solvation free energy of the complex is important, one can appreciate the potential speed-up afforded by reducing the number of poses that enter into the partition function calculation.

Figure 4 illustrates the level of sampling and the kinds of poses that contribute to the partition function for *o*-xylene. Poses with relative Boltzmann factors greater than 0.1 sample locally around the native pose. Including poses with Boltzmann factors between 0.01 and 0.1 brings in poses in which the ring is flipped in a nonsymmetry-related way. Poses with Boltzmann factors between 0.001 and 0.01 show even more extensive sampling. Poses with Boltzmann factors less than 0.001 contribute negligibly to the partition function for this system. Boltzmann factors of 0.001, 0.01, and 0.1 correspond roughly to relative energies of 2, 5, and 7 *kT*, respectively, above the global minimum.

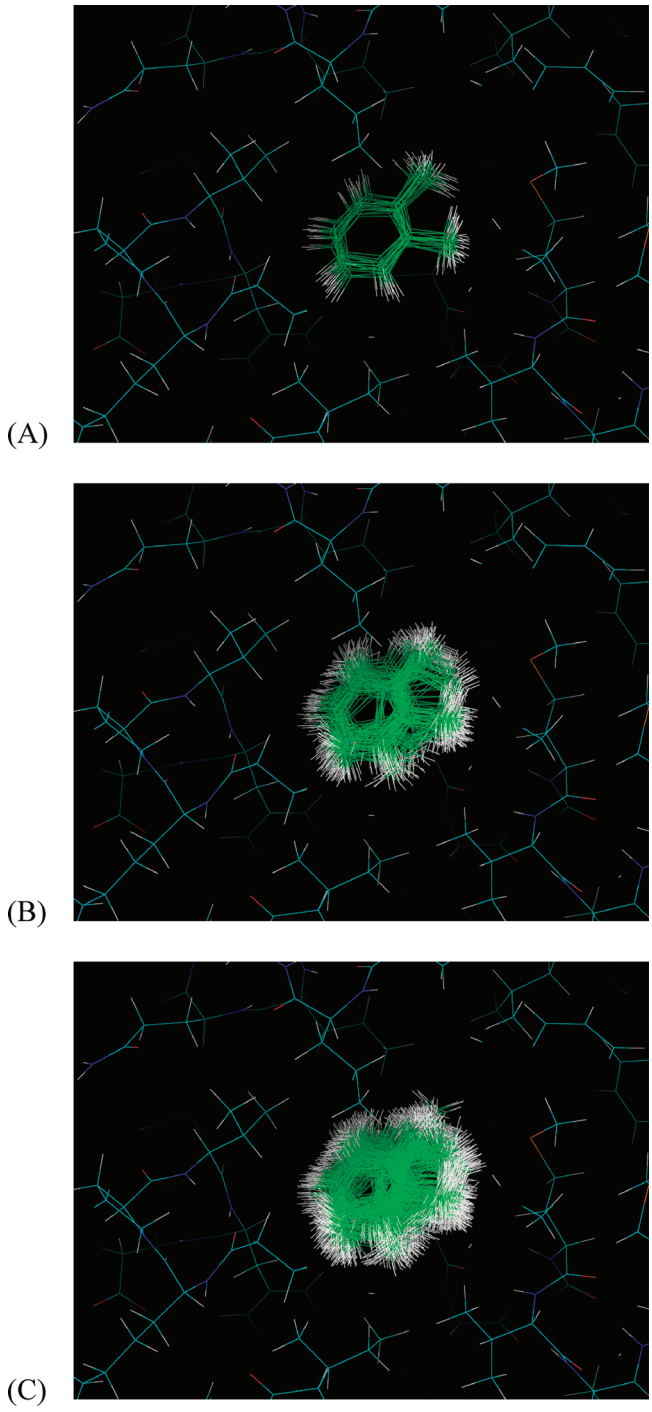


Figure 4. Overlay of poses of *o*-xylene whose relative Boltzmann factor is greater than (A) 0.1, (B) 0.01, and (C) 0.001. These correspond roughly to poses with energies of <2 *kT*, 5 *kT*, and 7 *kT*, respectively, with respect to the global minimum-energy pose.

Entropy. Given the partition function, one can readily calculate the average binding energy, which, when subtracted from the free energy, gives the configurational entropy of binding. It should be noted that what we refer to as the average energy has mixed into it solvation free energy effects and hence cannot be equated to the enthalpy obtained in an experimental measurement. The entropic terms for the two electrostatic models are nearly identical despite the differences in the average energies. This is because the partition functions are built up from the same exhaustive docking ensembles, and the

dx.doi.org/10.1021/jp212646s | *J. Phys. Chem. B* 2012, 116, 6872–6879

shape of the potential energy surface for the complex is not much changed for these largely nonpolar molecules as a function of the solute dielectric constant. Figure 5 is a scatter plot of the configurational entropy versus average energy. Aside from benzene, we see an anticorrelation between the entropy and average energy of binding with a Pearson correlation coefficient r^2 = 0.74. This anticorrelated behavior echoes that observed by Chen et al.[25] in their study of cyclodextrin host–guest complexes.

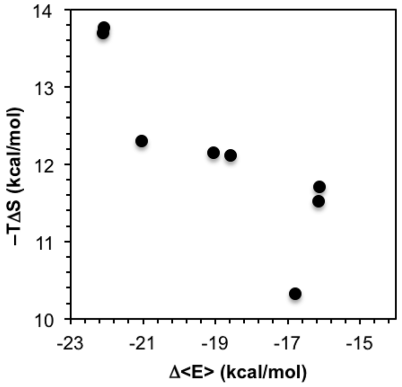


Figure 5. Scatter plot of entropy versus average energy.

For isobutylbenzene and *n*-butylbenzene, there are two sources of entropy loss upon binding. One is the reduction in translational and rotational entropy. The other is the reduction in torsional entropy. In an attempt to characterize the loss of torsional entropy, we carried out the same protocol as above except that a single rigid conformer of the two ligands was used. The energy-minimized bound conformation was used for exhaustive docking. Table 3 summarizes the free energies, enthalpies, and entropies for these two compounds using a single rigid ligand conformer. The difference in entropies between Tables 1 and 3 gives an estimate of the loss of torsional entropy. For electrostatic model 1, we observe a torsional $-T\Delta S$ of binding of 1.29 and 1.71 kcal/mol, respectively, for isobutylbenzene and *n*-butylbenzene. The results for electrostatic model 2 are similar. This is an entropic cost of about 0.6 kcal/mol per rotatable bond. Our torsional sampling is very local and ignores the possible reduction in the number of wells between the bound and free states. Hence, our calculated torsional entropic cost arises entirely from a steepening of the potential energy well and may be an underestimate. By comparison, Chang et al.[37] estimated losses in torsional entropy in the range of 0.6 to 1.0 kcal/mol per rotatable bond in their test system of HIV protease and amprenavir. Their estimate also includes a contribution from the reduction in number of energy wells. It appears that our estimates of the entropic costs of binding per rotatable bond are consistent with their findings using an independent method on a different system. However, it should be noted that decomposition of entropy into torsional and other terms is not entirely rigorous since the different degrees of freedom are coupled.[37]

Binding Cavity Definition. In the model system studied here, the L99A binding cavity is completely enclosed. One can therefore unambiguously define a ligand found in the cavity to be the bound state of the ligand. In general, binding cavities are partly open to the solvent, and there exists a continuum of states between a ligand that is clearly bound and one that is clearly in bulk water. In that situation, the size and location of the search box for exhaustive sampling is subject to judgment and is somewhat arbitrary. When the binding affinity is high and the region of favorable energy is fairly localized, the exact dimensions of the sampling box will not be too important as long as it contains the high-affinity regions. A box that spills over into "non-binding" regions will add terms to the summation in eq 2 that involve negligible Boltzmann factors. Greater sensitivity to the box definition can arise for weak binders for which low-energy poses are not well localized. The difficulty of precisely defining the bound state has been previously recognized and analyzed in depth.[33,38,39]

CONCLUSIONS

We have shown the feasibility of using exhaustive docking to build up a partition function for free energy calculations for the simple but still challenging model system provided by T4 lysozyme L99A. The results are competitive with those obtained from other methods that have used the same model system.[18−23] However, in its current stage, the method presented here is by no means ready for production runs for use with flexible ligands and/or binding sites. However, it is an encouraging first step in exploring the use of exhaustive docking as a viable sampling method for free energy estimation. We feel that several factors were important in obtaining reasonable results:

(1) Exhaustive docking ensures that all relevant ligand orientations are included in the free energy calculation with no bias or over/undersampling artifacts that may be present in stochastic methods.
(2) The bulk of the protein was kept fixed and identical for the different ligands in order to reduce noise from irrelevant regions while allowing limited flexibility close to the ligand to capture some of the energetics of conformational change induced by the ligand.
(3) For the more flexible ligands, local torsional sampling around the lowest-energy bound conformation approximated the steepening of the potential energy well in going from the free to the bound state.
(4) The solvation of the binding site is unaltered in the free and bound states, eliminating a major source of uncertainty in the more general case.
(5) The continuum solvation model properly treated the binding cavity as desolvated even in the apo state.

Table 3. Binding Free Energies and Components Using a Single Rigid Conformation for *i*-Butylbenzene and *n*-Butylbenzene[a]

	electrostatic model 1			electrostatic model 2		
	ΔG	$\Delta\langle E\rangle$	$-T\Delta S$	ΔG	$\Delta\langle E\rangle$	$-T\Delta S$
i-butylbenzene	−9.35 (0.20)	−21.83 (0.18)	12.48 (0.18)	−8.35 (0.20)	−20.82 (0.38)	12.46 (0.18)
n-butylbenzene	−9.47 (0.04)	−21.44 (0.05)	11.98 (0.05)	−8.56 (0.04)	−20.53 (0.09)	11.97 (0.05)

[a]All values are in kcal/mol. The calculated average energies and entropic contributions are shown as well.

dx.doi.org/10.1021/jp212646s | J. Phys. Chem. B 2012, 116, 6872–6879

Recently, Wlodek et al.[40] have described the use of an exhaustive conformational sampler with normal-mode analysis derived from Quasi-Newton energy minimization as an efficient way to estimate free energies of flexible ligands in solution. They also presented initial data on binding free energies as well. The combination of exhaustive docking in combination with exhaustive conformational search and QN normal-mode analysis is a future direction that may provide an efficient viable alternative to current methods for estimating binding free energies that are based on molecular dynamics or Monte Carlo simulations.

AUTHOR INFORMATION

Corresponding Author

*E-mail: enrico.purisima@nrc.gc.ca.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This is publication number 53160 of the National Research Council of Canada.

REFERENCES

(1) Gohlke, H.; Klebe, G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644.
(2) Lee, M. S.; Olson, M. A. *Biophys. J.* **2006**, *90*, 864.
(3) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150.
(4) Cui, Q.; Sulea, T.; Schrag, J. D.; Munger, C.; Hung, M.-N.; Naïm, M.; Cygler, M.; Purisima, E. O. *J. Mol. Biol.* **2008**, *379*, 787.
(5) Naïm, M.; Bhat, S.; Rankin, K. N.; Dennis, S.; Chowdhury, S. F.; Siddiqi, I.; Drabik, P.; Sulea, T.; Bayly, C.; Jakalian, A.; Purisima, E. O. *J. Chem. Inf. Model.* **2007**, *47*, 122.
(6) Sulea, T.; Hogues, H.; Purisima, E. O. *J. Comput.-Aided Mol. Des.* **2012**, online.
(7) Sulea, T.; Cui, Q.; Purisima, E. O. *J. Chem. Inf. Model.* **2011**, *51*, 2066.
(8) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.
(9) Wang, J.; Purisima, E. O. *J. Am. Chem. Soc.* **1996**, *118*, 995.
(10) Wang, J.; Szewczuk, Z.; Yue, S.-Y.; Tsuda, Y.; Konishi, Y.; Purisima, E. O. *J. Mol. Biol.* **1995**, *253*, 473.
(11) Head, M. S.; Given, J. A.; Gilson, M. K. *J. Phys. Chem. A* **1997**, *101*, 1609.
(12) Purisima, E. O. *J. Comput. Chem.* **1998**, *19*, 1494.
(13) Purisima, E. O.; Nilar, S. H. *J. Comput. Chem.* **1995**, *16*, 681.
(14) Salaniwal, S.; Manas, E. S.; Alvarez, J. C.; Unwalla, R. J. *Proteins Struct. Funct. Bioinf.* **2006**, *66*, 422.
(15) Case, D. A.; Cheatham, T. E. III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M. Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.
(16) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
(17) Morton, A.; Matthews, B. W. *Biochemistry* **1995**, *34*, 8576.
(18) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118.
(19) Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1255.
(20) Gallicchio, E.; Lapelosa, M.; Levy, R. M. *J. Chem. Theory Comput.* **2010**, *6*, 2961.
(21) Clark, M.; Meshkat, S.; Wiseman, J. S. *J. Chem. Inf. Model.* **2009**, *49*, 934.
(22) Irudayam, S. J.; Henchman, R. H. *J. Phys. Chem. B* **2009**, *113*, 5871.
(23) Carlsson, J.; Aqvist, J. *J. Phys. Chem. B* **2005**, *109*, 6448.
(24) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535.
(25) Chen, W.; Chang, C. E.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 3035.
(26) Williams, C.; Feher, M. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 39.
(27) McGann, M. *J. Chem. Inf. Model.* **2011**, *51*, 578.
(28) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. *Biopolymers* **2003**, *68*, 76.
(29) Sulea, T.; Purisima, E. O. The solvated interaction energy (SIE) method for scoring binding affinities. In *Computational Drug Discovery and Design*; Baron, R., Ed.; Humana Press (Springer Publishing Group): New York, 2012; p 295.
(30) Bhat, S.; Purisima, E. O. *Proteins Struct. Funct. Bioinf.* **2006**, *62*, 244.
(31) Chan, S. L.; Purisima, E. O. *J. Comput. Chem.* **1998**, *19*, 1268.
(32) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047.
(33) Gallicchio, E.; Levy, R. M. Recent theoretical and computational advances for modeling protein−ligand binding affinities. In *Advances in Protein Chemistry and Structural Biology*; Christo, C., Ed.; Academic Press: New York, 2011; Vol. *85*; p 27.
(34) Purisima, E. O.; Corbeil, C. R.; Sulea, T. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 373.
(35) Sulea, T.; Wanapun, D.; Dennis, S.; Purisima, E. O. *J. Phys. Chem. B* **2009**, *113*, 4511.
(36) Genheden, S.; Kongsted, J.; Söderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.* **2010**, *6*, 3558.
(37) Chang, C. E.; Chen, W.; Gilson, M. K. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1534.
(38) Luo, H.; Sharp, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99 10399.
(39) Mihailescu, M.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 23.
(40) Wlodek, S.; Skillman, A. G.; Nicholls, A. *J. Chem. Theory Comput.* **2010**, *6*, 2140.