Calculation of Configurational Entropy with a Boltzmann-Quasiharmonic Model: The Origin of High-Affinity Protein-Ligand Binding

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JUNE 2011

Impact Factor: 3.3 · DOI: 10.1021/jp111176x · Source: PubMed

2 AUTHORS:

Kyle Harpole
University of Pennsylvania
5 PUBLICATIONS **32** CITATIONS
SEE PROFILE

Kim A Sharp
University of Pennsylvania
176 PUBLICATIONS **18,645** CITATIONS
SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

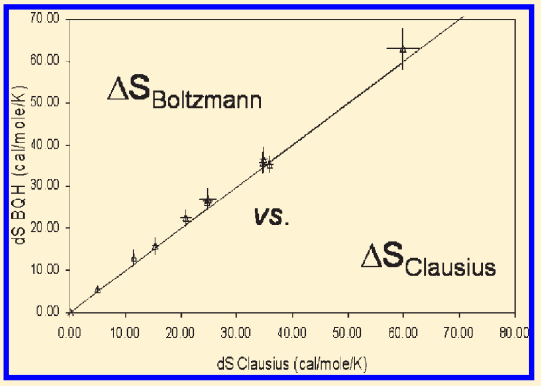
Calculation of Configurational Entropy with a Boltzmann−Quasiharmonic Model: The Origin of High-Affinity Protein−Ligand Binding

Kyle W. Harpole and Kim A. Sharp*

Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, United States

Supporting Information

ABSTRACT: Accurate assessment of configurational entropy remains a large challenge in biology. While many methods exist to calculate configurational entropy, there is a balance between accuracy and computational demands. Here we calculate ligand and protein conformational entropies using the Boltzmann−quasiharmonic (BQH) method, which treats the first-order entropy term by the Boltzmann expression for entropy while determining correlations using the quasiharmonic model. This method is tested by comparison with the exact Clausius expression for entropy on a range of test molecules ranging from small ligands to a protein. Using the BQH method, we then analyze the rotational and translational (R/T) entropy change upon ligand binding for five protein complexes to explore the origins of extremely tight affinity. The results suggest that in these systems such affinity is achieved by a combination of simultaneously maintaining good protein−ligand contacts while allowing significant residual R/T motion of the ligand through suitable protein motions.



INTRODUCTION

The binding affinity of a ligand for its protein partner depends on the balance between intermolecular interactions between the binding partners, mutual desolvation of the binding partners, and configurational entropy changes. The entropy change arises from three sources: changes in conformational fluctuations of the ligand (L), changes in conformational fluctuations of the protein (P), and reduction of the translational and rotational (R/T) freedom of the ligand with respect to the protein (or vice versa). Changes in internal conformational fluctuations of the ligand have long been a focus of drug development, where, for example, cyclization or rigidification of the ligand is designed to reduce the ligand's entropy loss upon binding. However, this strategy does not always increase binding affinity because other enthalpic and entropic contributions are affected.[1,2] Protein fluctuations contribute to the entropy and have been elucidated in particular by modern NMR methods[3−5] and changes upon ligand binding examined.[6,7] NMR studies have shown that for Calmodulin−peptide binding changes in the residual protein entropy upon ligand binding play a controlling role in ligand specificity.[8,9]

While the sources of the configurational entropy change upon binding are known, obtaining precise estimates for specific binding reactions is a major challenge. One approach has been to calculate free energy of binding (ΔG^{bind}) directly through, for example, calculation of thermodynamic work. Methods such as free energy perturbation (FEP) and thermodynamic integration have been used for accurate calculation of binding free energy changes due to ligand changes.[10−14] These are best suited to obtaining relative free energies, typically from small ligand changes. A related method, the linear interaction free energy method,[15] estimates absolute free energies. Since these methods only provide free energies, additional calculations such as *T*-derivatives must be performed if one is specifically interested in entropy contributions.[16,17] Other methods include calculating the potential of mean force (PMF) for bringing the ligand into the binding site[18] and calculating the thermodynamic work of "decoupling" the ligand from the protein into a "restrained" reference unbound state with known translational/rotational entropy.[19−23] Free energy and thermodynamic work based methods for entropy calculation include contributions from fluctuations in solute and solvent contributions, and correlations between them. Solvent effects have previously been shown to affect solute entropies[24] and protein−ligand binding entropies through direct and indirect effects.[18]

The other approach to calculating solute configurational entropy changes is from the solute coordinate fluctuations. The advantage of this approach is that one can in principle identify individual contributions, e.g., from R/T type motions of

Received: November 23, 2010
Revised: June 15, 2011

the ligand with respect to the protein, from internal conformational changes in protein and ligand, i.e., torsional, angle bend, and bond stretch (denoted here as the internal, or TAB, conformational changes). One can also localize changes in entropy to parts of the protein[25,26] to obtain structural and mechanistic insight into what controls binding affinity. Of particular interest is the translational/rotational contribution ($\Delta S_{R/T}$). Attempts to develop accurate estimates for protein−ligand binding have been numerous.[27−38] The major problem has been in estimating the precise amount of residual translational/rotational entropy of the bound ligand. Many treatments of binding, particularly empirical screening functions and calculations of relative binding for closely related ligand series, have been forced to assume that $\Delta S_{R/T}$ is either negligible or constant.[39] To go beyond this requires accurate representation of the motions of the complex, and one must account for correlations in R/T motions and internal fluctuations of the two binding partners. Progress in this area relies on having a firm statistical mechanical treatment of the total binding free energy (ΔG^{bind}) and total binding entropy change (ΔS^{bind}).[40−44]

In principle, the configurational entropy can be obtained from Boltzmann's equation

$$S = -k \int p(\mathbf{q}) \ln(p(\mathbf{q}))\, d\mathbf{q} \quad (1)$$

by integration of the multidimensional probability density function (pdf) p($\mathbf{q}$) over the appropriate coordinate space $\mathbf{q}$. In practice the integration is of such high dimensionality that sampling requirements are prodigious, and direct integration is impractical.[45] Recently Li et al.[46] found that even evaluation of three-dimensional pdf's became inaccurate for more than 80 degrees of freedom. Thus the major effort has been toward developing different approximations to the full Boltzmann equation and assessing their accuracy in multiatomic and ligand−protein systems. It has not proven easy to test the accuracy of a particular approximation; one must first ensure that precision of the numerical implementation (i.e the sampling and integration) of the method is good enough to test the approximation itself. Thus calculation of entropy changes due to ligand−protein binding remains challenging.

An early and significant advance is this area was the development of the quasiharmonic (QH) approximation.[47−52] Here the coordinate fluctuation probability distribution is approximated by a multidimensional Gaussian. The entropy is then obtained from the determinant or Eigenvalues of the coordinate fluctuation covariance matrix. This method has been frequently used to calculate macromolecular entropies and was applied with some success to calculate ligand−protein binding entropies and free energies.[41,53,54]

Other methods applied to peptides and proteins include the hypothetical scanning approach[55] and the nearest neighbor method.[56] Gilson et al. have published a detailed and careful series of studies on the calculation of configurational entropy in which they have explored techniques such as mining minima,[54,57−59] mode scanning,[60] and various approximations to the Boltzmann equation.[26,61,62] One focus has been on improving the correlation term, for example by accumulation of second and higher order probability distributions, such as in mutual information expansion (MIE).[25,46,51,62,63]

Inclusion of increasingly higher order distribution terms will in principle increase the accuracy of the entropy calculation, but sampling requirements rapidly become severe. The attraction of the original quasiharmonic method is that although it includes a second-order correlation term, this only requires computation of covariances and not complete two-dimensional probability distributions. However, a significant drawback is that the first-order term (i.e., the diagonal term of the fluctuation matrix), which represents the entropy contribution without correlation between different types of motions, also assumes a Gaussian probability distribution.[64] For cases where the ligand is fairly rigid in both the bound and free states, this a reasonable approximation,[41,54] but this assumption is unrealistic for many cases and it can lead to significant errors.[64−67] For example, it is likely that torsional angle distributions of a group or side chain exploring multiple rotamers will not even be unimodal let alone Gaussian. A second limitation of the original QH method is that it only includes pairwise correlations and not higher order correlations. A significant effort has been made to ameliorate these two limitations of QH method. One approach is to add corrections for anharmonicity and higher order correlations. Rojas et al. explicitly included a cubic anharmonic correction.[64] Grubmüller et al. used adaptive anisotropic kernels and minimum information methods to achieve accurate entropies for systems comprising up to 10s of atoms.[45,68] Baron et al. have included explicit corrections to the QH method for anharmonicity and higher order correlations and used this to obtain well-converged conformational entropies for a number of systems, including peptides and lipids.[66,67,69,70]

Di Nola et al. addressed the limitations of the QH method by factoring out the first-order entropy term (the diagonal term) from the fluctuation covariance matrix.[71] The first term now represents the QH entropy neglecting all correlations. The normalized covariance matrix in the second term is now the QH correction due to correlations. They then *replaced* the first-order term with the exact uncorrelated entropy expression, a sum of terms of the form $\int p(q_i) \ln(p(q_i))\, dq_i$, where the integration is over one-dimensional pdf's for each coordinate q_i. Thus in the leading entropy term there is no assumption about Gaussian distributions or unimodality. The attraction of their approach is that both terms remain rather tractable to evaluate: the first term because only N one-dimensional probability distributions are needed, where N is the number of degrees of freedom, and the second term because although the number of second-order terms scales as N^2, each term requires calculation of just a covariance, not full two-dimensional or higher distribution functions. This should be contrasted with other approaches to extending the original QH method in which the QH approximation is used for both the first-order term (the leading diagonal terms in the covariance fluctuation matrix) and the pairwise second-order term (the off-diagonal terms), with subsequent corrections to both first-order and correlation terms. Here we refer to the method of Di nola et al. as the Boltzmann−quasiHarmonic (BQH) method to highlight the use of the Boltzmann expression for the leading term. We apply their method to different ligand and protein ligand systems and assess the accuracy of the method in these systems.

A problem with assessing the accuracy of any approximation to the full Boltzmann expression, eq 1, is that the actual entropy for molecules consisting of more than a few atoms is not known analytically. One approach is to compare different kinds of approximation to eq 1. For example Li et al. compared the straight QH method in a system known to be close to harmonic, with the MIE method. Another approach is to compare the effects

dx.doi.org/10.1021/jp111176x |J. Phys. Chem. B XXXX, XXX, 000–000

Table 1. Test Molecules

compound	protein partner	PDB[a]	atoms	N_{rotate}[b]	sampling time (ns)	frames sampled
O_2	na	na	2	0	1	10 000
butane	na	na	14	3	4	40 000
biotin	streptavidin	1STP	31	5	5	50 000
digoxin	antibody 26−10	1IGI	62	6	4	40 000
trimethoprim	DHFR	na	21	7	4	40 000
methotrexate	DHFR	na	33	12	4	40 000
PT523	DHFR	1OHK	43	13	4	40 000
NADPH	DHFR	1OHK	75	18	4	40 000
ouabain	antibody 40−50	1IBG	85	12	4	40 000
HIV-1 peptide (PEPTAPPEE)	TGG101 peptide binding protein	1M4P	128	27	10	100 000
peptide P_{10}	profilin	1AWI	143	10	10	100 000
benzamidine	trypsin	3PTB	18	2	8	80 000
carboxyl arabinotyl 2,5 phosphate (CAP)	RUBISCO	1RBO	30	12	8	80 000
streptavidin	streptavidin	1STP	1744	513	23	230 000

[a] PDB entry from which starting structure was taken. na: Not applicable. [b] Number of rotatable bonds (defined as single, nonring bonds).

of increasing levels of correction to the QH method.[45,46,66−70] An alternative approach is to compare to a method of entropy calculation not based on eq 1. For example Grubmüller et al.[45,68] used thermodynamic integration (TI) as their "gold standard", obtaining first the free energy and then the absolute entropy by subtracting the energy/enthalpy term. The TI method is analytically exact, in the sense that the equations for entropy themselves involve no approximation. Only the numerical evaluation is approximate. We use a similar strategy here, using the analytically exact Clausius equation for entropy to benchmark our calculations. For small to medium size molecules this enables us to put a rather small bound on the error from the approximation in the correlation term. We then explore the feasibility of applying the method to full protein−ligand binding entropy changes, in terms of the sampling and convergence requirements. Finally, we apply the method to study the rotational/translational entropy change in protein−ligand systems in order to gain insight into the origin of high-affinity protein−ligand binding.

THEORY AND METHODS

Molecular Dynamics Simulations. Molecular dynamics (MD) simulations were performed using the CHARMM molecular modeling package.[72,73] Starting coordinates of the complex, protein, or ligand alone were taken from the Protein Databank (PDB) where appropriate (Table 1). Hydrogen atoms were added and alternate polar hydrogen positions searched to optimize H-bonds where appropriate. Parameters for proteins were taken from the standard CHARMM27 force field.[73] For ligands not in this parameter set, bonded parameters and van der Waals parameters were transferred from chemically homologous groups or moieties. Atomic charges were obtained using the electronegativity neutralization method, Q_{equil}. This method, based on that of Rappe and Goddard,[74] has been developed by us to rapidly obtain charges for a wide range of chemical groups.[75] The method was specifically designed to accurately reproduce experimental solvation energies using the Poisson−Boltzmann/SAS method used in previous binding free energy calculations and in hybrid MD simulations.[76,77] After assignment of parameters and building of hydrogens the structures were exhaustively minimized using the ABNR minimizer.

Molecular dynamics simulations were performed with a time step of 1 fs and a nonbonded cutoff of 14 Å using the shift method. Simulations of ligands alone were performed with a distant-dependent dielectric of $4r$ to model electrostatic screening from solvent. While this is not a particularly realistic model, it was chosen for two reasons. (a) To compare configurational entropy calculation methods, it is necessary to reduce uncertainty due to sampling. Simulations with $\varepsilon = 4r$ are sufficiently rapid to ensure that configurational entropy estimates converge in a reasonable amount of time. (b) The comparison of two different methods for calculating ligand configurational entropy involves taking temperature differences (see below). In contrast to explicit solvent models, an implicit solvent model can be kept T-invariant. Thus the only explicit degrees of freedom that can give rise to an entropy *change* are the TAB and R/T coordinates; i.e.,the T-dependence reveals just protein or ligand internal conformational entropy changes, and R/T changes for complexes. Simulations were run with Langevin dynamics, using an atomic friction coefficient of 70 ps^{-1} scaled by the fractional surface accessibility of the atom. The Langevin dynamics approach was selected because it models solvent fluctuation effects missing in an implicit solvent model, and it improves conformational sampling. It also provides excellent temperature control. Simulations were equilibrated for 1 ns and sampled for a period of time commensurate with the entropy convergence behavior (Table 1). Frames were saved every 100 fs for entropy calculations. Simulations involving proteins and protein−ligand complexes were performed using the Poisson−Boltzmann−solvent accessible surface area (PBSA) method to model solvent effects on electrostatics and hydrophobicity, respectively. The smooth permittivity implementation of the PBSA model was used,[78] as described previously.[76,77] This PBSA model is fully integrated into CHARMM. The solvent and protein dielectrics were $\varepsilon = 80$ and $\varepsilon = 1$, respectively, and a hydrophobic area coefficient of $\gamma = 5$ (cal/mol)/$Å^2$ was used. For simulations of proteins in which only torsional motions were allowed, bond lengths and angles were fixed using the SHAKE method.[79]

Calculation of Configurational Entropy. *Boltzmann−Quasiharmonic Method.* For calculation of configurational entropies, molecular coordinates were converted to a nonredundant internal coordinate representation of torsion angles, bond angles,

dx.doi.org/10.1021/jp111176x |J. Phys. Chem. B XXXX, XXX, 000–000

and bond lengths (referred to here as TAB or internal conformational coordinates) represented in Z-matrix form. The method for calculation of dS^{conf} is based on that of Di Nola et al. for estimation of the internal entropy of a protein.[71] In the QH model the entropy contribution from fluctuations is given by[48]

$$\Delta S^{conf}/k = \frac{1}{2}\ln((2\pi e)^n|{\sigma_{ij}}^2|) \quad (2)$$

where k is the Boltzmann constant and ${\sigma_{ij}}^2$ is the coordinate fluctuation covariance matrix, whose elements are given by

$${\sigma_{ij}}^2 = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle \quad (3)$$

where q_i is the instantaneous value of the ith internal coordinate and $\langle\ \rangle$ indicates an ensemble average. The QH approximation is by definition inaccurate if fluctuations are very non-Gaussian. This is often the case for internal coordinate representations where there is an isomerization between two or three distinct rotamers and the torsional fluctuations would be poorly represented as a single Gaussian.[65] In the QH model the determinant in eq 2 can be factored into a product of the diagonal elements, D, and the determinant of the normalized fluctuation covariance (or correlation coefficient) matrix C_{ij} as

$$|{\sigma_{ij}}^2| = \prod_i^n {\sigma_{ii}}^2 |C_{ij}| = D|C_{ij}| \quad (4)$$

with elements $C_{ij} = {\sigma_{ij}}^2/({\sigma_{ii}}^2{\sigma_{jj}}^2)^{1/2}$. Now $^1/_2 \ln(2\pi e)^n D$ represents the contribution of the fluctuations to entropy without accounting for correlation between motions, while $^1/_2 \ln|C_{ij}|$ represents the correction due to correlations. In the QH model both contributions are estimated with the Gaussian approximation. As Di Nola et al. have shown,[71] a more accurate representation is obtained by replacing the diagonal term with the exact Boltzmann expression for the entropy in the absence of correlations, $S/k = \sum \int p(q_j) \ln p(q_j)\, dq_j$, where $p(q_j)$ is the probability distribution of coordinate q_j and the sum is over all coordinates. So now, assuming that probability distributions of all the internal coordinates are represented by the usual histogram binning method, we have

$$\Delta S^{conf}/k = -\sum_q^n \sum_i^m \Delta q {p_i}^q \ln {p_i}^q + {}^1/_2 \ln |C_{ij}| \quad (5)$$

where ${p^q}_i$ is the probability density for internal coordinate q in bin i. The inner sum is over the m bins of equal size Δq into which internal coordinate q's range is divided. Here the reference state for entropy is that of a uniform distribution over q's range, $m\Delta q$, but the choice of this reference state does not affect entropy changes. The outer sum is over all the internal coordinates. For protein−ligand complexes, the internal coordinates of eqs 2 and 3 also include three translational and three rotational coordinates (referred to here as R/T coordinates) describing the relative motion of the ligand with respect to the protein. Here the center of mass and principle moments of inertia of the protein and ligand were used to define their relative position and orientation. In C_{ij} the bond torsion angles, ϕ, were represented in complex form as $z = e^{i\phi}$ to handle the discontinuity in ϕ (the modulo 2π problem).[51] The rotational/translation entropy in the unbound reference state is computed as $k \ln(1660.8\pi^2)$, where 1660 Å^3 is the reference volume at 1 M in the length units used here. Errors were estimated from the standard deviation in entropy from batch means, dividing the total data into 10 batches.

The difference in configurational entropy upon temperature change or binding was obtained by evaluating eq 5 for the two cases and taking the difference. Individual components of the entropy were obtained from eq 5 as follows. TAB and R/T contributions were obtained from the plnp term of eq 5 by summing over the contributions from pdf's for each type of degree of freedom. The contribution from all correlations within and between these degrees of freedom comes from the second term in eq 5. Householders elimination was used to calculate the determinant of C_{ij}.[80] Correlation components were obtained by analyzing the contribution from each subblock of the normalized fluctuation covariance matrix C_{ij} corresponding to T, A, B, and R/T motions, or pairwise combinations thereof, as described by Chang et al.[54] Briefly, the contribution is obtained from the difference in correlation entropy with and without inclusion of a particular set of correlation terms .

In summary, the DiNola et al.'s method, which we refer to here as the BQH method, only uses a Gaussian approximation to estimate the effect of correlations between motions: Formally, the total correlation between motions of coordinates i and j is approximated by the correlation in second moments of their distributions at the pairwise level. A more exact treatment of the correlations would require the sampling and convergence of three-dimensional and higher dimension phase integrals, which currently is intractable for proteins due to the combinatorial explosion in the number of such pdf's.[61]

Clausius Method. To assess the accuracy of the BQH method, we also calculated the entropy of test systems by an alternative method that does not depend on an approximation to the Boltzmann expression for entropy, eq 1. From the thermodynamic identities

$$C_p = \frac{dH}{dT} = T\frac{dS}{dT} \quad (6)$$

the equation for the change in entropy due to a finite change in temperature is

$$\Delta S = \frac{\Delta H}{T_{vH}} \quad (7)$$

where T_{vH} is a particular form of average temperature, the van't Hoff average, given by

$$T_{vH} = \frac{T_2 - T_1}{\ln(T_2) - \ln(T_1)} \quad (8)$$

(see Supporting Information). T_1 and T_2 are the initial and final temperatures, respectively. Equation 7 is a finite difference form of the classic Clausius expression for entropy $\Delta S = \int dq/T$, where q is the heat absorbed. If the heat capacity of the system is constant over the temperature interval, then it is an exact expression for the entropy when the van't Hoff average temperature given by eq 8 is used. If not, it remains an extremely accurate numerical approximation, as shown in the Supporting Information. Thus, unlike the QH or BQH models, it starts from an analytically exact expression for the entropy. The change in enthalpy, ΔH, is obtained from the difference in mean potential energies from two CHARMM MD simulations at temperatures of T_1 and T_2. (The difference here between ΔE and ΔH is

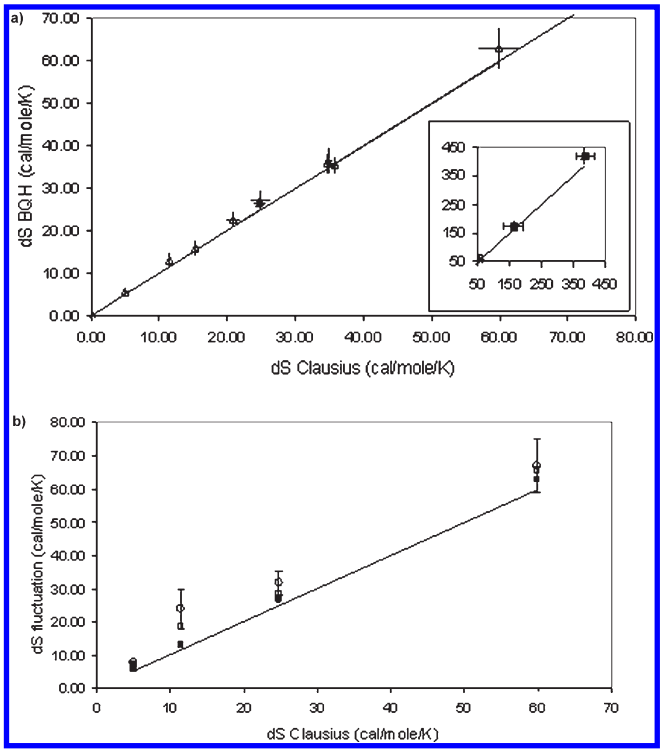


Figure 1. (a) Comparison of temperature-induced entropy changes calculated from the Clausius expression (eq 7) and the BQH method (eq 5), including torsion, angle, bond stretch (TAB) terms, and all six sets of intraset correlations, T−T, A−A, B−B, T−A, T−B, A−B. (Δ). The two data points for streptavidin (inset, ■) include all contributions (upper point) or torsion and torsion correlations alone (lower point). (b) Comparison of entropy changes calculated from the Clausius expression with the BQH method (■), the BQH method neglecting all correlations (□), and the quasiharmonic method (eq 2; ○).

negligible.). By using a *T*-independent implicit solvent model, the entropy change in eq 7 depends only on changes in TAB and R/T entropy, so one can compare this value with the corresponding entropy contributions calculated from fluctuations in molecular coordinates using the BQH approximation. By this means we can assess the accuracy of the BQH approximation.

Binning of Internal Coordinate. When the distribution of a continuous internal coordinate q is represented in discrete form as a binned histogram, correct choice of bin width Δq is essential to ensure that the calculated entropy is independent of the discretization. At one extreme, using a single bin of width q_{range}, the entropy would by definition attain its maximum theoretical reference value of $k \ln(q_{\text{range}})$, blind to any variation in the actual distribution. Conversely, division into increasingly narrow bins would result in a large number of empty bins. This would indicate a spuriously precise localization of conformations in phase space with an entropy value that decreases without limit. This problem has been satisfactorily addressed by Edholm and Berendsen.[81] Between extremes of bin width, the calculated entropy will plateau over some range of widths that depends on the natural range, distribution, and resolution of that coordinate. The suggested bin width given by Edholm and Berendsen is proportional to $1/N^{1/3}$, where N is the number of samples. This is similar to the standard formula in statistics for the optimal bin width of $\Delta q = 2.49\sigma/N^{1/3}$,[82] where σ is the standard deviation of the data. The practical solution to this issue is to start with a bin

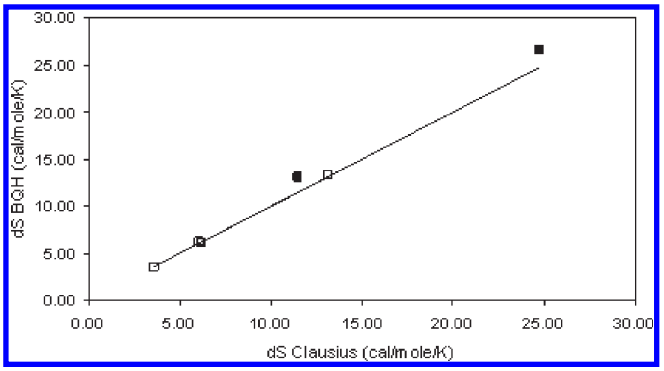


Figure 2. Comparison of ligand entropy changes calculated from the Clausius and BQH methods for free ligands (TAB + all correlations; ■) and ligands bound to protein (TAB + R/T + correlations; □).

width several fold narrower than that suggested by either formula, to accumulate the data, and to aggregate in increasingly wider bins to find the plateau region, as done previously.[66,70] This can be accomplished with minimal extra cost after the initial coordinate analysis. In tests on typical protein/ligand coordinate fluctuations, as indicated in the results below, we have found a bin width of 3−5° to be appropriate for angular quantities, and 0.1 Å for length/position variables.

For uniform bin width, the entropy for each coordinate q is most conveniently obtained by

$$S^q/k = \ln(N\Delta q) - \frac{1}{N}\sum_i^m n_i \ln n_i \qquad (9)$$

where N is the total number of samples, n_i is the number of counts in the ith bin of m bins total, and Δq is the bin width.

■ RESULTS

To test the accuracy of the Boltzmann−quasiharmonic approximation, the configurational entropy change of a free ligand due to a change in temperature was calculated two ways: by the BQH method (eq 5) and the Clausius method (eq 7). Simulation temperatures of 280 and 320 K were used, with ΔT = 40 K. Eleven ligands ranging in size from 2 to 143 atoms were examined along with one protein, streptavidin (1744 atoms, PDB code 1STP; see Table 1.) The entropy includes the TAB contributions and all correlations within and between these three sets of degrees of freedom. Figure 1a shows that the agreement between the two methods is excellent for all of the ligands, with all data falling within the error bars of the unit slope line. The inset with reduced scale shows the data for the two cases with the largest entropy, PEPTAPPEE and streptavidin. The agreement is also excellent for the protein, which has an order of magnitude more atoms. A linear regression for the entire set gave an R^2 value of 0.997. Figure 1b shows, for comparison, the entropy change for four ligands (butane, biotin, digoxin, and PEPTAPPEE) calculated using either the BQH method (eq 5), the full QH method (eq 2), or the BQH method neglecting the correlation term (i.e., using just the $p \ln p$ term of eq 5). The QH values fall above the line with significant disagreement with the Clausius method. It is not surprising that the QH model, which assumes a Gaussian distribution of fluctuations, would overestimate entropies, since, for a given fluctuational variance, the Gaussian distribution is the one with the maximum entropy.[50,70] The magnitude of this overestimation has been documented in studies of solute entropy

dx.doi.org/10.1021/jp111176x |J. Phys. Chem. B XXXX, XXX, 000–000

Table 2. Contributions to Temperature-Induced Entropy Change[a]

		ΔS ((cal/mol)/K)[c]				ΔS_{norm} [d] (k)		
	N_{free} [b]	bond	angle	torsion	correlation	bond	angle	torsion
biotin	87	3.9	4.0	10.7	−5.4	0.98	1.04	2.87
digoxin	180	8.1	7.9	11.5	−1.6	1.00	0.99	1.47
PEPTAPPEE	378	16.8	16.74	26.8	−2.63	0.99	1.00	1.61
peptide P_{10}	423	9.7	10.0	16.9	−0.10	1.00	1.03	1.76

[a] For peptide ligands, $T_1 = 280$ K, $T_2 = 300$ K, for other ligands $T_2 = 320$ K. [b] Internal TAB degrees of freedom. [c] Bond, angle, torsion: Uncorrelated contributions given by the $p \ln p$ term of eq 5. Correlation: From all correlations within and between sets of degrees of freedom, i.e., T−T, A−A, B−B, T−A, T−B, and A−B given by the second term of eq 5. [d] Entropy change normalized by $0.5kN_d \ln(T_2/T_1)$, where N_d is number of bond, angle or torsion degrees of freedom.

Table 3. Entropy Change Components for Biotin[a]

	temperature change 280→320 K			binding to streptavidin			
	torsion	angle	bond	torsion	angle	bond	trans/rot.
component							
ΔS_{uncorr} [b]	10.7	4.0	3.9	−7.0	−0.9	−1.4	−30.8
correlation[c]							
torsion	−6.1	0.7	0.0	3.0	−1.3	0.01	−2.0
angle		0.01	0.01		−0.06	−0.14	0.0
bond			0.0			−0.07	0.0
trans/rot.							−3.7

[a] Entropies in units of (cal/mol)/K or eu. [b] ΔS_{uncorr}: Uncorrelated torsion, angle, bond (TAB) contributions given by the $p \ln p$ term of eq 5. [c] Correlation components within and between the indicated sets of degrees of freedom obtained from the corresponding blocks of the normalized fluctuation covariance matrix C_{ij} of eq 5 using the method of Chang et al.[54]

on other systems,[66,67] where it can produce 10−200% overestimates. It is also clear that there is no simple linear relationship; i.e., the error from the Gaussian distribution approximation is not a constant, correctable factor. In addition, neglecting the correlation term also results in clearly inferior agreement. This is also expected, since correlations between motions would have the effect of reducing the magnitude of the total entropy.[61,83]

In Figure 2 the entropy change as a function of temperature was calculated for four ligands, biotin, digoxin, benzamidine, and CAP, while bound to their protein receptor. Entropy changes for two ligands, biotin and digoxin in the unbound state (taken from Figure 1a), are shown for comparison. In all cases the protein was held rigid, so only the ligand TAB and R/T degrees of freedom can contribute to the entropy change. This was done to test the accuracy of the BQH approximation when R/T configurational fluctuations were included while preventing the large effects of protein TAB contributions from swamping the calculation. The possibility arises that QH approximation to the correlation term may be more accurate for some degrees of freedom than others. In Figure 1, the only correlation contributions are from TAB internal conformation fluctuations. Figure 2 includes the translational and rotational libration contributions, and correlations within and between R/T and TAB motions. The agreement

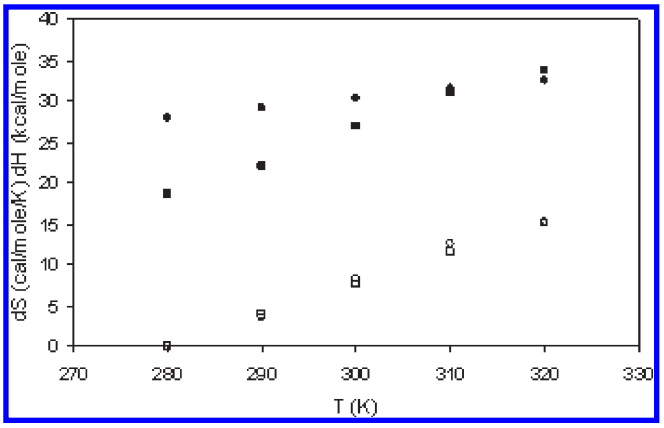


Figure 3. Effect of temperature increment on the entropy calculation for trimethprim. Mean ligand potential energy (●), mean ligand entropy (■), change in entropy from Clausius (□), and BQH methods relative to 280 K (○). Entropies include TAB and all correlation terms.

between the Clausius and BQH methods, however, is equally good if not better. It should be emphasized the strategy of isolating different contributions to entropy by freezing out certain degrees of freedom does not imply that these entropy contributions are additive. Indeed, the correlation terms in Tables 2 and 3 show this is not the case. The BQH equation is an approximation, and the goal in Figures 1 and 2 is to test how accurate this approximation is by isolating specific contributions and comparing with the Clausius equation.

Inspection of Figure 1 shows that the BQH values are slightly but systematically higher than the Clausius equation values with the linear regression giving a slope of 1.05 ± 0.03. Various sources for this disagreement were examined. We first examined the temperature increment, to see if the finite difference approximation to the Clausius equation was a significant source of error. Figure 3 shows the configurational entropy of trimethoprim calculated by the BQH method at five temperatures ranging from 280 to 320 K in 10° increments, and the difference in this entropy relative to that at 280 K. The plot also shows the mean energy ($E \approx H$) of trimethoprim, and the difference in entropy relative to T = 280 K calculated from enthalpy using the Clausius method. The agreement between the two entropy calculations is almost unchanged for ΔT values ranging from 10 to 50 K. The finite difference approximation to the Clausius equation should degrade for larger ΔT (but will remain small—see Supporting Information). Conversely, the relative precision of the BQH method should increase at larger ΔT, since the difference in fluctuations is larger. Since the two methods agree to the same extent throughout, we conclude that errors due to the finite value of ΔT are not the source of the discrepancy.

Another possibility is that the selected bin size for the internal coordinates may lead to inaccuracies in the BQH method. The effect of internal coordinate bin size is shown in Figure 4. For digoxin simulated at 280 and 320 K the bond torsion, angle, and length fluctuations were collected using from 50 to 360 bins. For angular variables this gave bin widths ranging from 7 to 1°, while for length variables the width ranged from 0.7 to 0.1 Å. The entropy at each temperature shows a moderate decrease as the bin width is initially narrowed, followed by slower decrease up to $N = 360$. With finer bins ($N > 360$) the downward slope increases progressively (data not shown). The difference in entropy, however, is almost independent of bin width over this range,

dx.doi.org/10.1021/jp111176x |J. Phys. Chem. B XXXX, XXX, 000–000

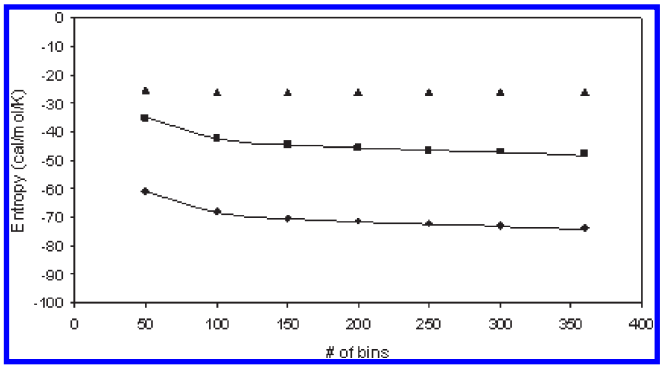


Figure 4. Entropy change for digoxin as a function of bin size. Mean entropy at 280 (●), 320 (■), and the difference (▲). Solid lines are fits to the function $S = S_0 - AN + B/N^2$.[81] Entropies include TAB and all correlation terms.

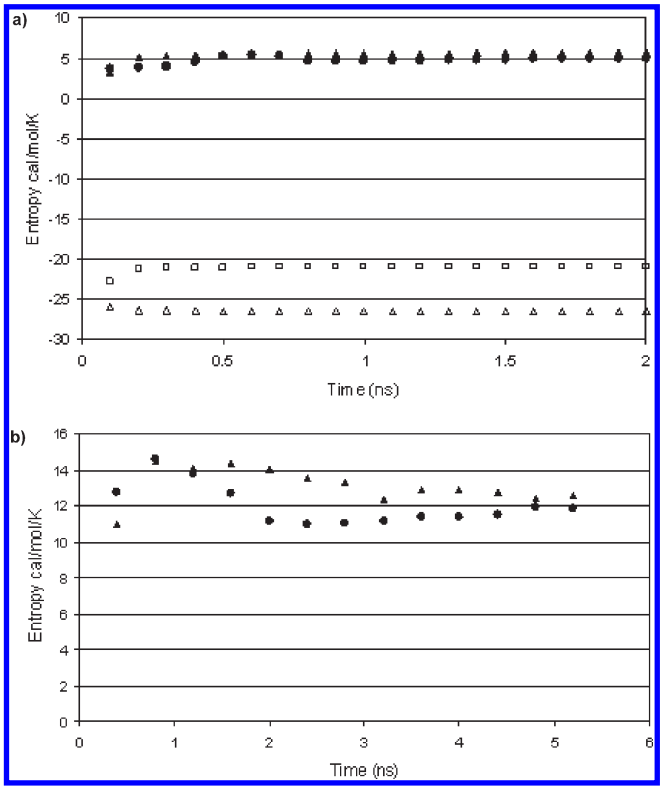


Figure 5. Convergence of entropy change for (a) butane and (b) biotin. Mean entropy computed by BQH method at 280 (Δ) and at 320 K (□) and the difference (▲). Entropy difference from the Clausius equation (●). Entropies include TAB and all correlation terms.

varying from −25.3 to −25.9 eu (1 (cal/mol)/K = 1 entropy unit or eu.). The variation in entropy at both temperatures is well-fit by the functional form suggested by Edholm and Berendsen, $S = S_0 - AN + B/N^2$,[81] where S_0, A, and B are constants (Figure 4, solid lines). The best fit line for both data sets corresponds to A = 0.015 eu and B = 150 eu for both temperatures, and S_0 = −43 and −69 eu at T = 280 and 320 K, respectively. The difference in the bin width independent parameter ΔS_0 = 26.0 eu, providing a highly precise and bin-independent way to estimate entropy differences. Under the conditions used for Figure 1, entropy changes are insensitive to bin width. This rules out binning effects as a source of the small difference between BQH and Clausius entropy estimates.

The last effect we considered was different rates of convergence for BQH and Clausius method estimates. Figure 5 shows the convergence behavior of the entropy differences due to a 280−320 K temperature change for butane and biotin. In both cases the estimates from the two methods converge at similar rates, and the BQH entropy is systematically larger in magnitude throughout, except for the 1 ns equilibration period at the beginning. Similar convergence behavior is seen for the other ligands. Together, these three results imply that the 5% larger entropy estimate from the BQH method is a real effect stemming from the approximate treatment of the correlation term.

Overall, these results demonstrate that the BQH method provides a precise method for calculating configurational entropy changes for a variety of ligands. We are now in a position to examine different components of the configurational entropy, which can be easily extracted from the BQH method but not from the Clausius method. Details are given in Table 2 for four ligands. For all cases, the torsional contribution is the largest. This is not unexpected, as torsional motions are the softest type of deformation. The correlation term is also a significant and highly variable contribution to the total entropy. Bond stretch terms are usually the smallest contribution, again as expected because they are the stiffest mode of deformation. The exception is the PEPTPPEE peptide, where stretch and angle terms are comparable.

Bond stretch and angle terms are typically represented by harmonic terms in modern force fields, and CHARMM is no exception. The configurational entropy change of a 1-D harmonic oscillator due to a temperature change from T_1 to T_2 is given by $\Delta S^{\mathrm{harm}} = 0.5k \ln(T_2/T_1)$.[84] One can thus define a normalized entropy contribution, relative to that from harmonic motion as

$$\Delta S^{\mathrm{norm}}/k = \Delta S^i/(0.5N_{\mathrm{d}}^i k \ln(T_2/T_1)) \qquad (10)$$

in Boltzmann units, where the superscript i refers to total bond, angle, or torsion contributions, respectively, and N_{d}^i is the number of degrees of freedom of bond, angle, or torsion coordinates respectively, in the molecule. A value of 1 indicates perfect harmonic behavior as a function of temperature. These normalized bond, angle, and torsion entropies are shown in Table 2. Entropy changes from bond stretch and angle follow closely what one would expect from harmonic motions, while the torsional term is 50−100% greater than expected from harmonic motions.

Table 3 breaks down the motional correlations of the entropy of biotin into those between individual components: torsion−torsion, torsion−bond, and so on. For the free ligand the temperature change increases torsional, bending, and stretching entropies, the torsional term increasing most. The effect of the correlations between different motions is to reduce the entropy of those motions,[83] giving rise to a net negative contribution from the correlation term (Table 2). The dominant contribution is due to correlations between different torsional motions. The other terms are rather small, especially those involving only bends and stretches. To first order, only intratorsional correlations matter. For biotin binding to streptavidin, the first-order (uncorrelated) entropy changes are negative, with a large contribution of more than −30 eu from translational/rotational (relative to free unrestricted rotation and translation at a standard state of 1 M). Correlations between the six translational/rotational

dx.doi.org/10.1021/jp111176x |J. Phys. Chem. B XXXX, XXX, 000–000

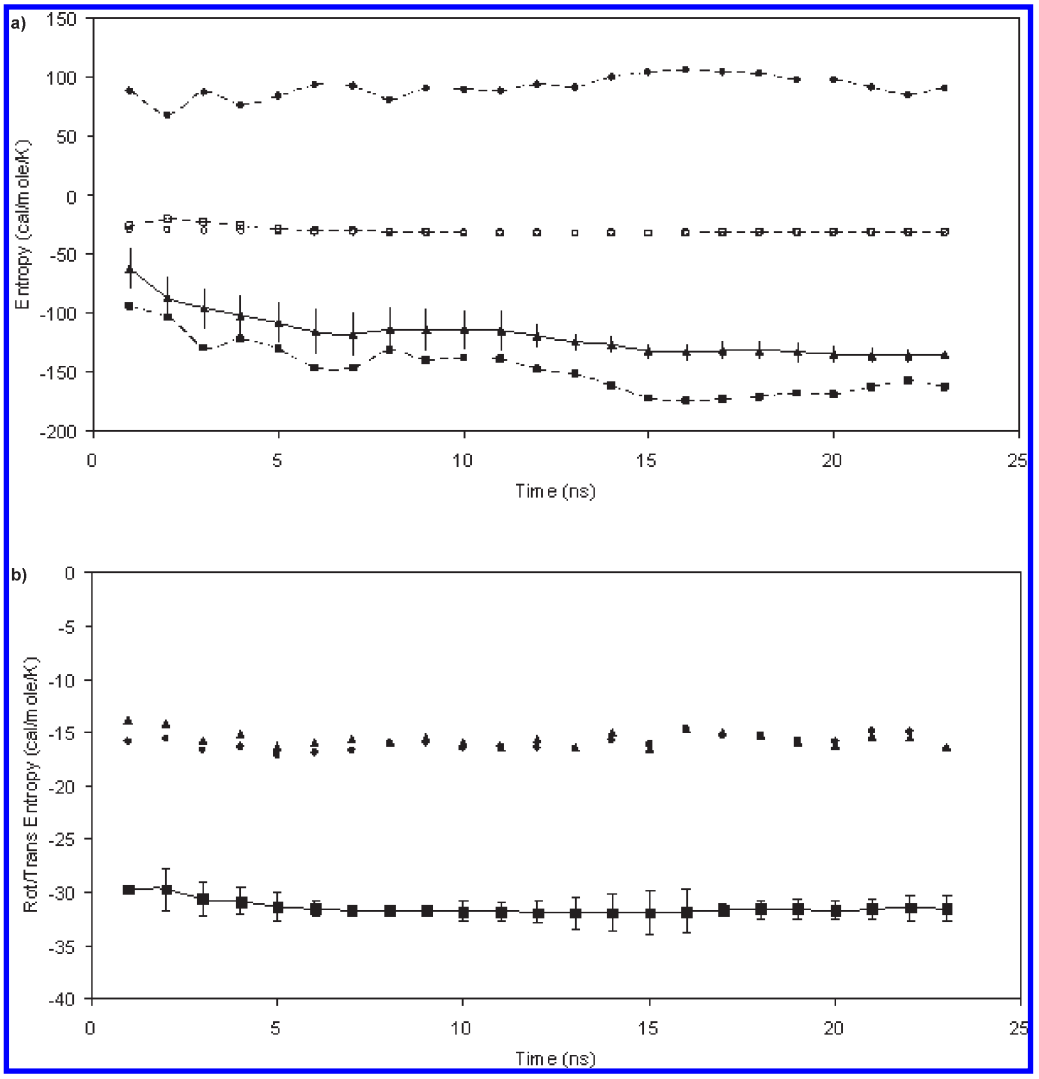


Figure 6. Convergence behavior of entropy components for biotin binding to streptavidin. (a) Protein torsions (■), ligand torsions (□), translation/rotation (○), total correlation (●), and total entropy change (TAB + R/T + all correlations; ▲). (b) Batch average values of rotational (●) and translational (▲) contributions. Cumulative average rotational plus translational contribution (■).

motions add a further entropy penalty of about −4 eu. Again intratorsional correlations offset the net negative first-order ΔS from reduction in torsional fluctuations. The only significant correlation between translation/libration of the ligand and internal motions is with the torsional portion. Correlations involving bending and stretching motions are small and to first order may be neglected. The small amount of correlation involving bend/stretch motions is consistent with their near harmonic behavior seen in Table 2: If individual bond bending and stretching motions were significantly coupled to other motions, one might expect them to become anharmonic.

Entropy changes were calculated using the BQH method for binding of biotin to strepavidin by simulating the ligand and protein separately, simulating the protein−ligand complex, and taking the difference in total entropy and entropy components. The data are shown in Figure 6a as a function of simulation length in order to assess the rate of convergence. Contributions from biotin torsions and translation/rotation converged rapidly, being effectively flat minimized after 5 ns. Bond stretch and angle contributions from both protein and ligand are not shown, but converge at a similar rate. Torsion contributions from the protein, however, converge considerably slower, as does the correlation term, and neither are fully converged at the end of 23 ns of simulation. This is not unexpected, as torsional motions are softer and undergo larger fluctuations, and combined with the large number of atoms in the protein, more sampling is required. Correlations are also expected to converge more slowly as they involve pairs of motions.[61,66] For a flexible peptide undergoing folding−refolding transitions it was found that the correlation term converged about half as rapidly as the quasiharmonic term alone.[66] It should be noted, however, that for the biotin−streptavidin complex the correlation and torsion terms are themselves highly anticorrelated, so the total entropy change undergoes smaller fluctuations than either of these subcomponents. Component analysis of the correlations (as in Tables 2 and 3) indicate that the bulk of the correlation arises from intratorsion correlations and cross-correlations involving torsions. In this study we were particularly interested in the rotational and translational entropy contributions to binding, so Figure 6b shows a more detailed view of the time course of these components. The net

dx.doi.org/10.1021/jp111176x |J. Phys. Chem. B XXXX, XXX, 000–000

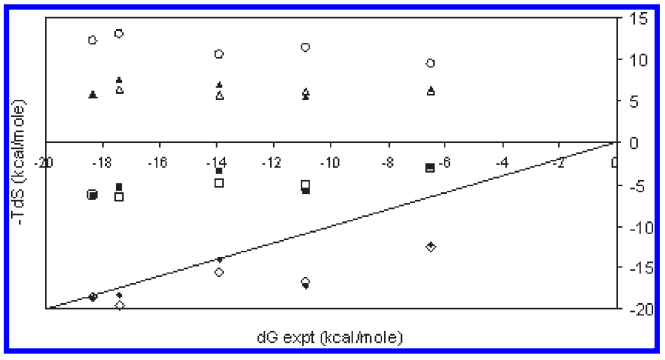


Figure 7. Rotation/translation entropy contributions to protein−ligand binding for the following: biotin−streptavidin, dixogin−antibody 2610, carboxyl arabinotyl 2,5 phosphate−RUBISCO, benzyl mercaptopropionate−carboxypeptidase, and benzamidine−trypsin. Entropies include all correlations with and between R and T degrees of freedom. Data are plotted as free energy contributions, $-T\,\Delta S$ at 300 K vs the experimental binding free energy. From binding to rigid protein, $\Delta S_{R/T}$(rigid), ○; from recovery due to protein flexibility, $\Delta S_{R/T}$(recover), □, ■; from net $\Delta S_{R/T}$ (rigid plus recovery), Δ, ▲; from the difference, $\Delta G^{proxy} \approx -T(-\Delta S_{R/T}(\text{rigid}) + \Delta S_{R/T}(\text{recover}))$ (see text for details), ♦, ◇. Unfilled and filled symbols are for 4 and 8 ns simulations, respectively. The $x = y$ line is shown for reference only.

R/T entropy undergoes fluctuations of less than 2−3 eu and is well-converged at the end of the simulation.

Using the BQH method, we examined the R/T contribution to the binding reaction L + P → LP for several protein ligand complexes. $\Delta S_{R/T}$ was computed by performing MD simulations on the protein−ligand complex, analyzing the R/T motion of the ligand in the bound complex to obtain the entropy, and subtracting the R/T entropy of the unbound ligand reference state at 1 M. The results are shown in Figure 7. To check on the convergence in these systems, data from 4 and 8 ns simulations were compared. We observed only minor differences. The net loss of R/T entropy was almost independent of the experimental binding free energy ($R^2 = 0.09$). The contribution of changes in R/T motion was then broken down into two steps, "loss" and "recovery". Loss corresponds to binding of the flexible ligand to the protein while the protein is held fixed in the conformation seen in crystal structure of the complex, denoted as L + P → LP(rigid). Recovery corresponds to relaxation of the protein with the ligand bound, denoted as LP(rigid) → LP. For the first step, the R/T entropy relative to its unbound state reference state at 1 M was obtained from a simulation of the ligand fluctuating in the rigid protein pocket. $\Delta S_{R/T}$ for the second step was obtained from the difference between the translational/rotational fluctuations of the ligand in the complex with and without constraints on the protein. The ligand was completely unconstrained in all the simulations. The rationale for breaking down the R/T entropy contribution into loss and recovery components this way is explained more fully in the discussion below. These two contributions are shown in Figure 7. The first contribution, $\Delta S_{R/T}$(rigid) correlates quite well with ΔG_{expt} ($R^2 = 0.77$). The second, recovery component $\Delta S_{R/T}$(recover) correlates somewhat better ($R^2 = 0.88$, respectively).

DISCUSSION

We have examined the accuracy of the BQH approximation for configurational entropy changes by comparing calculated values to an essentially exact thermodynamic method for a range of test systems. The BQH approximation does well even for relatively large, flexible systems containing 10^2−10^3 atoms such as peptide ligands or an entire protein. Both internal conformation and rotation/translation components are obtained precisely. Moreover, the small but systematic overestimation by the BQH method is in the expected direction and allows us to put a bound of 5% on the approximation for the systems studied here. The overestimation appears to be independent of size, as the overestimation is consistent from small rigid ligands through to a protein.

In this regard it is interesting to consider why the approximation works as well as it does. Considering the second-moment approximation, it would be exact for the pairwise term if the fluctuations were Gaussian. One can argue that it remains a very reasonable approximation even if the actual distribution of a particular coordinate is non-Gaussian. This is because the extent of correlation is similar in the second and higher moments, as shown by Baron et al.[66,70] As for the neglect of triplet and higher correlations, the results suggest that, at least for the molecules examined here, the contribution of the higher order correlations to *entropy* is small. Rojas et al.[64] applied a cubic correction to the quasiharmonic method and found that the resulting values for butane entropies were very close to the exact values. If most of this correction is accounting for the non-Gaussian 1-D probability distributions (which in the BQH method are treated explicitly by the leading Boltzmann term), this would also imply a small contribution from higher order terms to the correlation term. Explicit consideration of triplet and higher terms in other systems also supports this.[61,63]

In all of the molecules examined here, the dominant entropy term comes from torsional motions. Bond stretch and angle contributions are very close to those expected from stiff harmonic motion. For biotin−streptavidin binding, the overall distribution of entropy between torsions, bends, stretches, rotation/translation, and correlation term is similar to that seen in other systems.[26,54] Torsional contributions dominate, and there is a substantial contribution from R/T motions. For the correlation term, torsion−torsion and torsion−R/T terms are the most significant. The small contribution from intrabond stretch and intrabond angle correlations is consistent with the near harmonic motions of these degrees of freedom. For the R/T contribution to entropy, it is interesting to compare the values here for a very high affinity ligand−protein binding to that of peptide−protein binding.[26] This contribution was shown to be considerably smaller for peptide binding. Biotin binds in a deep cleft, and protein contact with the rigid half of the ligand necessarily restricts R/T motion. In contrast the peptide binds in a surface cleft, and with many more torsional degrees of freedom, the protein contacts act primarily to restrict internal torsion motions and not overall R/T motion. Of course with regard to the net effect on binding, the relevant factor is the large ligand entropy loss; however it is partitioned between internal and R/T motions.

The convergence behavior of any solute entropy calculation method is difficult to predict a priori. It depends on the method, treatment of solvent, the size, and flexibility of the molecule(s), and what component of the entropy is being examined. In previous studies of biological systems with explicit solvent time requirements have ranged from 2.5 ns simulations of lipids with frequent snapshot sampling[69] to >100 ns for flexible peptide−protein binding.[26] Simulations as long as 1 μs have been required for a peptide undergoing large folding−unfolding transitions.[66] Methods that evaluate full two-dimensional and higher order pdf's will require longer simulations than those

dx.doi.org/10.1021/jp111176x |J. Phys. Chem. B XXXX, XXX, 000–000

requiring one-dimensional pdf's. Moreover, explicit solvent models, with more explicit degrees of freedom and a more granular energy landscape probably require longer simulations than implicit solvent models. Whatever method is used, multiple independent trajectories can help with convergence particularly for entropy calculations since this reduces the statistical dependence of samples.[26,69] For the drug-size ligands examined here using the BQH method, which requires only 1-D pdf's, and Langevin dynamics with implicit solvent, adequate convergence of total solute conformational entropy changes is achieved with ca. 4 ns of molecular dynamics simulations. Peptide ligands of >5−10 amino acids (aa's) take somewhat longer, up to 10 ns, depending on their rigidity. Estimating the TAB conformational entropy change for ligand binding to a protein takes substantially longer, even with a second-order method such as BQH. We examined the convergence for the net entropy change for biotin−streptavidin binding. Streptavidin is a small 133 aa protein which has a relatively rigid closed β-barrel structure, so one would expect its convergence characteristics to be good as far as protein simulations go. While R/T and ligand TAB entropy changes converge rapidly, it would take considerably longer than 20 ns for the total correlation and protein torsion terms to converge. For all of the molecules examined here, the rate limiting terms for convergence of the total entropy change are those from torsional motions and correlations of those motions with themselves.

The R/T entropy change is of particular interest for protein−ligand binding, since estimates for it vary so widely. For the systems studied here, this term converges well enough to be studied with sub-10 ns simulations. The R/T contributions might be expected to converge slowly, since they correspond to large-scale "low-frequency" motions. However, for the tight binding protein−ligand complexes studied here there is a highly restricted range of motion. In the five complexes studied, root mean square (rms) translational motions are ≈0.4−0.8 Å, and rms librations are ≈8−20°. This limited phase space is sampled relatively quickly.

Tight binding of a ligand by a protein involves conflicting requirements thermodynamically: As the intermolecular interactions become increasingly favorable, binding is tighter, but the ligand becomes more restricted in its residual translational/rotational motion, and probably many ligand and protein internal motions are more restricted too. Indeed the general trend is that as ligands increase in size (and hence in number of possible interactions), there are diminishing returns in incremental affinity.[85,86] The trade-off between configurational entropy and interaction terms is treated explicitly in statistical mechanical treatments of binding (see for example refs 40−44 and discussions therein) and has been documented in calculations of ligand cyclodextrin binding.[87]

Assuming that R/T entropy can be obtained reliably using the BQH method, we looked at this contribution to binding in five protein−ligand complexes: biotin−streptavidin, dixogin−antibody 2610, carboxyl arabinotyl 2,5 phosphate−RUBISCO, benzyl mercaptopropionate−carboxypeptidase, and benzamidine−trypsin. The first three are very tight binding, relatively speaking, in that their affinities lie considerably above the "Kuntz−Kollman" curve.[85] The other two have typical binding affinity relative to this curve. The net $\Delta S_{R/T}$ contribution to binding is almost completely uncorrelated with the experimental binding free energy, ΔG^{bind} in these systems (Figure 7, $R^2 = 0.09$), and at first sight $\Delta S_{R/T}$ provides no insights into the origin of tight binding. The binding reaction L + P → LP was also broken down into to two steps: (1) binding of the flexible ligand to the protein while holding the latter in its bound conformation L + P → LP(rigid); (2) relaxation of the protein with the ligand bound LP(rigid) → LP. The first contribution, $\Delta S_{R/T}$(rigid), represents an upper bound on the loss of R/T entropy. It also effectively measures the goodness of fit of the ligand in the protein pocket. A loose fit will result in a lower magnitude of $\Delta S_{R/T}$(rigid). We thus reasoned that, at least in relatively high affinity binding, $\Delta S_{R/T}$(rigid) is a loose readout or proxy for the net contribution of intermolecular interactions (including solvent) to binding: Hence a large loss indicates better interactions, i.e., a more negative $T\ \Delta S_{R/T}$(rigid) indicates a more negative ΔG(interaction). However, if this was the only factor in play, improved interactions would not result in substantial gain in binding, since they are partially offset by the R/T restriction. The recovery term measures how much residual R/T entropy is gained because protein flexibility allows for additional net ligand motion in the pocket. If the protein could allow this libration/translation motion while still maintaining good intermolecular contacts, this would be all gain as far as net binding free energy is concerned. It is known that receptor flexibility can make large contributions to binding entropy changes and that such motions are highly correlated to binding,[67] providing a mechanism for achieving both good contacts plus good residual R/T motion to achieve high-affinity binding. This effect should show up as a large negative value for the *difference* in the two $\Delta S_{R/T}$ components above. Thus their contribution to the binding free energy is

$$\Delta G^{proxy} \approx -T(-\Delta S_{R/T}(\text{rigid}) + \Delta S_{R/T}(\text{recover})) \qquad (11)$$

where, to reiterate, a contribution equal to the first entropy loss term is counted as *favorable*, since, in tight binding complexes, it reflects formation of favorable intermolecular interactions. When we look at these two "components" of $\Delta S_{R/T}$, we find that they do correlate with binding affinity, especially the recovery term. Moreover, their difference, as given by eq 11, correlates well with the experimental affinity ($R^2 = 0.83$). Of course, while the trends are clear, absolute comparison with experimental free energies would require enthalpic and solvent free energy contributions to be included.

It should be noted that splitting up the net R/T entropy contribution to binding into two steps, the first involving binding to a rigid protein does not say anything about the additivity of the different contributions to TAB and R/T entropy. Nor is it a unique decomposition. Rather, as in the pathway analysis of free energy changes,[88] a carefully chosen decomposition can provide "mechanistic or structural" insight into changes in thermodynamic quantities. Applied here, this analysis of R/T motions indicates that one strategy proteins employ for high-affinity binding is to combine good intermolecular contacts with good residual freedom of ligand motion in the complex. Although this analysis does not include explicit accounting of the ligand and protein internal entropy contributions, nor the contribution from solvent entropy, it is clear that increasing residual R/T motion while maintaining good contacts must entail increased internal motion: For the rigid ligand cases this must come from the protein. This again acts to reduce the configurational entropy penalty of binding. It is enlightening to compare this strategy with that of proteins that must mediate promiscuous but high-affinity binding. In a recent study of entropy in such systems using the QH method, Chang et al.[89] found that the high affinity

dx.doi.org/10.1021/jp111176x |J. Phys. Chem. B XXXX, XXX, 000–000

binding partners manifest alternative favorable contacts. It is apparent that in this case too the entropic penalty of binding was ameliorated while maintaining good contacts.

CONCLUSIONS

A major problem in testing methods for computing configurational entropy is that the true entropy in a biological system of any complexity is unknown. One approach is to compare different approximations to the $S = -k\int p(\mathbf{q}) \ln(p(\mathbf{q}))\, d\mathbf{q}$ integral or examine the convergence as successive terms are added to some expansion approximation.[61] Either way, the methods are all using the same coordinate fluctuations. In this work we compared the approximate statistical mechanical BQH method to the Clausius *T*-variation thermodynamic method. Here one is comparing entropies calculated from coordinates vs energetic quantities. Although ultimately both depend on the coordinates, they do so in very different ways, so the methods are more independent, giving greater confidence to any agreement. The Clausius method was used for calibration, not because we are interested in temperature effects per se in protein−ligand binding. The BQH was then applied to entropy changes in protein binding, and to component analysis as neither can be done with the Clausius method. Despite this, the Clausius method could be used to validate any approximate coordinate fluctuation method for estimating configurational entropy.

ASSOCIATED CONTENT

Supporting Information. Text giving details on the Clausius method for computing entropy changes. This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: sharpk@mail.med.upenn.edu. Tel.: 215-573-3506.

ACKNOWLEDGMENT

We thank Josh Wand and Eric Manas for many discussions about entropy over the years.

REFERENCES

(1) DeLorbe, J. E.; Clements, J. H.; Teresk, M. G.; Benfield, A. P.; Plake, H. R.; Millspaugh, L. E.; Martin, S. F. *J. Am. Chem. Soc.* **2009**, *131*, 16758.
(2) Ward, J. M.; Gorenstein, N. M.; Tian, J.; Martin, S. F.; Post, C. B. *J. Am. Chem. Soc.* **2010**, *132*, 1058.
(3) Akke, M.; Bruschweiler, R.; Palmer, A. *J. Am. Chem. Soc.* **1993**, *115*, 9832.
(4) Yang, D.; Mok, Y.; Forman-Kay, J.; Farrow, N.; Kay, L. *J. Mol. Biol.* **1997**, *272*, 790.
(5) Li, Z.; Raychaudhuri, S.; Wand, A. J. *Protein Sci.* **1996**, *5*, 2647.
(6) Lee, A. L.; Kinnear, S. A.; Wand, A. J. *Nat. Struct. Biol.* **2000**, *7*, 72.
(7) Kranz, J. K.; Flynn, P. F.; Fuentes, E. J.; Wand, A. J. *Biochemistry* **2002**, *41*, 2599.
(8) Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. *Nat. Chem. Biol.* **2010**, *6*, 352.
(9) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. *Nature* **2007**, *448*, 325.
(10) Lybrand, T.; McCammon, J. A.; Wipf, G. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 833.
(11) Beveridge, D.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.
(12) Bash, P.; Singh, U. C.; Langridge, R.; Kollman, P. *Science* **1987**, *236*, 564.
(13) Bash, P.; Singh, C.; Brown, F.; Langridge, R.; Kollman, P. *Science* **1987**, *235*, 574.
(14) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742.
(15) Zhou, R.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. *J. Phys. Chem. B* **2001**, *105*, 10388.
(16) Fleischmann, S.; Brooks, C. *J. Chem. Phys.* **1987**, *87*, 3029.
(17) Peter, C.; Oostenbrink, C.; van Dorp, C.; van Gunsteren, W. *J. Chem. Phys.* **2004**, *120*, 2651.
(18) Baron, R.; Setny, P.; McCammon, J. A. *J. Am. Chem. Soc.* **2010**, *132*, 12091.
(19) Hermans, J.; Shankar, S. *Isr. J. Chem.* **1986**, *27*, 225.
(20) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670.
(21) Woo, H.-J.; Roux, B. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825.
(22) Wang, J.; Deng, Y.; Roux, B. *Biophys. J.* **2006**, *91*, 2798.
(23) Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1255.
(24) Baron, R.; Bakowies, D.; Van Gunsteren, W. F. *J. Pept. Sci.* **2005**, *11*, 74.
(25) Li, D.; Meng, D.; Bruschweiler, R. *J. Am. Chem. Soc.* **2009**, *131*, 14610.
(26) Killian, B. J.; Yudenfreund Kravitz, J.; Somani, S.; Dasgupta, P.; Pang, Y.-P.; Gilson, M. K. *J. Mol. Biol.* **2009**, *389*, 315.
(27) Steinberg, I.; Scheraga, H. *J. Biol. Chem.* **1963**, *238*, 172.
(28) Finkelstein, A. V.; Janin, J. *Protein Eng.* **1989**, *3*, 1.
(29) Janin, J. *Proteins-Struct., Funct., Genet.* **1995**, *24*, R1.
(30) Janin, J. *Proteins* **1996**, *21*, 30.
(31) Karplus, M.; Janin, J. *Protein Eng.* **1999**, *12*, 185.
(32) Yu, Y. B.; Privalov, P. L.; Hodges, R. S. *Biophys. J.* **2001**, *81*, 1632.
(33) Amzel, L. M. *Proteins* **1997**, *28*, 144.
(34) Brady, G. P.; Sharp, K. A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 215.
(35) Murphy, K. P.; Xie, D.; Thompson, K.; Amzel, M.; Freire, E. *Proteins* **1994**, *18*, 63.
(36) Mammen, M.; Shakhnovich, E.; Deutch, J.; Whitesides, G. *J. Org. Chem.* **1998**, *63*, 3821.
(37) Singh, N.; Warshel, A. *Proteins* **2010**, *78*, 1724.
(38) Singh, N.; Warshel, A. *Proteins* **2010**, *78*, 1705.
(39) Bohm, H. J. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309.
(40) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047.
(41) Luo, H.; Sharp, K. A. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 10399.
(42) Swanson, J. M.; Henchman, R.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67.
(43) Sharp, K. Potential functions for virtual screening and ligand binding calculations: Some theoretical considerations. In *Virtual Screening in Drug Discovery*; Alvarez, J., Shoichet, B., Eds.; CRC Press: Boca Raton, FL, 2005; Chapter 9.
(44) Zhou, H.-X.; Gilson, M. *Chem. Rev.* **2009**, *109*, 4092.
(45) Hensen, U.; Grubmüller, H.; Lange, O. F. *Phys. Rev. E* **2009**, *80*, 011913.
(46) Li, D.-W.; Showalter, S.; Bruschweiler, R. *J. Phys. Chem. B* **2010**, *114*, 16036.
(47) Go, N.; Scheraga, H. *J. Chem. Phys.* **1969**, *51*, 4751.
(48) Karplus, M.; Kushick, J. *Macromolecules* **1981**, *14*, 325.
(49) Levy, R.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules* **1984**, *17*, 1370.
(50) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617.
(51) Wang, J.; Bruschweiler, R. *J. Chem. Theory Comput.* **2006**, *2*, 18.
(52) Schafer, H.; Mark, A.; van Gunsteren, W. *J. Chem. Phys.* **2000**, *113*, 7809.
(53) Carlsson, J.; Aqvist, J. *J. Phys. Chem. B* **2005**, *109*, 6448.
(54) Chang, C. A.; Chen, W.; Gilson, M. K. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1534.

(55) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Phys.* **2006**, *125*, 024905.
(56) Hnizdo, V.; Darian, E.; Fedorowicz, A.; Demchuk, E.; Li, S.; Singh, H. *J. Comput. Chem.* **2007**, *28*, 655.
(57) Head, M. S.; Given, J. A.; Gilson, M. K. *J. Phys. Chem. A* **1997**, *101*, 1609.
(58) David, L.; Luo, R.; Gilson, M. K. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 157.
(59) Chen, K.; Gilson, M.; Webb, S.; Potter, M. *J. Chem. Theory Comput.* **2010**, *6*, 3540.
(60) Chang, C.-E.; Potter, M. J.; Gilson, M. K. *J. Phys. Chem. B* **2003**, *107*, 1048.
(61) Killian, B. J.; Kravitz, J.; Gilson, M. K. *J. Chem. Phys.* **2007**, *127*, 024107.
(62) Hnizdo, V.; Tan, J.; Killian, B. J.; Gilson, M. K. *J. Comput. Chem.* **2008**, *29*, 1605.
(63) Somani, S.; Killian, B. J.; Gilson, M. K. *J. Chem. Phys.* **2009**, *130*, 134102.
(64) Rojas, O.; Levy, R. M.; Szabo, A. *J. Chem. Phys.* **1986**, *85*, 1037.
(65) Chang, C.-E.; Chen, W.; Gilson, M. K. *J. Chem. Theory Comput.* **2005**, *1*, 1017.
(66) Baron, R.; Hunenberger, P. H.; McCammon, J. A. *J. Chem. Theory Comput.* **2009**, *5*, 3150.
(67) Baron, R.; McCammon, J. A. *ChemPhysChem* **2008**, *9*, 983.
(68) Hensen, U.; Lange, O. F.; Grubmüller, H. *PLoS One* **2010**, *5*, e9179.
(69) Baron, R.; deVries, A.; Hunenberger, P. H.; van Gunsteren, W. *J. Phys. Chem. B* **2006**, *110*, 15602.
(70) Baron, R.; Van Gunsteren, W. F.; Hünenberger, P. H. *Trends Phys. Chem.* **2006**, *11*, 87.
(71) Di Nola, A.; Berendsen, H.; Edholm, O. *Mol. Phys.* **1984**, *17*, 2044.
(72) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(73) MacKerell, A. D.; Brooks, B.; Brooks, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley & Sons: Chichester, U.K., 1998; Vol. 1, p 271.
(74) Rappe, A.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
(75) Yang, Q.; Sharp, K. A. *J. Chem. Theory. Comput.* **2006**, *2*, 1152.
(76) Prabhu, N. V.; Zhu, P.-J.; Sharp, K. A. *J. Comput. Chem.* **2004**, *25*, 2049.
(77) Prabhu, N.; Panda, M.; Yang, Q.; Sharp, K. *J. Comput. Chem.* **2008**, *29*, 1113.
(78) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608.
(79) Van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.
(80) Press, W.; Flannery, B.; Teukolsky, S.; Vetterling, W. *Numerical Recipes*; Cambridge University Press: Cambridge, U.K., 1986.
(81) Edholm, O.; Berendsen, H. *Mol. Phys.* **1984**, *51*, 1011.
(82) Scott, D. W. *Biometrika* **1979**, *66*, 605.
(83) Matsuda, H. *Phys. Rev. E* **2000**, *62*, 3096.
(84) Hill, T. *An Introduction to Statistical Thermodynamics*; Dover: New York, 1986.
(85) Kuntz, I. D.; Chen, K.; Sharp, K.; Kollman, P. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9997.
(86) Brooijmans, N.; Sharp, K. A.; Kuntz, I. D. *Proteins* **2002**, *48*, 645.
(87) Chen, W.; Chang, C.-E.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 3035.
(88) Boresch, S.; Karplus, M. *J. Mol. Biol.* **1995**, *254*, 801.
(89) Chang, C. A.; McLaughlin, W.; Baron, R.; Wang, W.; Andrew, M. J. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 7456.