Published in final edited form as:

Phys Chem Chem Phys. ; 24(10): 6037–6052. doi:10.1039/d1cp05075c.

Developing end-point methods for absolute binding free energy calculation using the Boltzmann-quasiharmonic model†

Lauren Wickstrom[a], **Emilio Gallicchio**[b,c,d], **Lieyang Chen**[c,d,e], **Tom Kurtzman**[c,d,e], **Nanjie Deng**[f]

[a]Borough of Manhattan Community College, The City University of New York, Department of Science, New York, New York, USA

[b]Department of Chemistry, Brooklyn College, The City University of New York, Brooklyn, New York, USA

[c]PhD Program in Chemistry, Graduate Center of the City University of New York, New York, USA

[d]PhD Program in Biochemistry, Graduate Center of the City University of New York, New York, USA

[e]Department of Chemistry, Lehman College, The City University of New York, Bronx, New York, USA

[f]Department of Chemistry and Physical Sciences, Pace University, New York, New York, USA

Abstract

Understanding the physical forces underlying receptor–ligand binding requires robust methods for analyzing the binding thermodynamics. In end-point binding free energy methods the binding free energy is naturally decomposable into physically intuitive contributions such as the solvation free energy and configurational entropy that can provide insights. Here we present a new end-point method called EE-BQH (Effective Energy-Boltzmann-Quasiharmonic) which combines the Boltzmann-Quasiharmonic model for configurational entropy with different solvation free energy methods, such as the continuum solvent PBSA model and the integral equation-based 3D-RISM, to estimate the absolute binding free energy. We compare EE-BQH with other treatments of configurational entropy such as Quasiharmonic models in internal coordinates (QHIC) and in Cartesian coordinates (QHCC), and Normal Mode analysis (NMA), by testing them on the octa acids host–guest complexes from the SAMPL8 blind challenge. The accuracies in the calculated absolute binding free energies strongly depend on the configurational entropy and solvation free energy methods used. QHIC and BQH yield the best agreements with the established potential of mean force (PMF) estimates, with R^2 of ~0.7 and mean unsigned error of ~1.7 kcal mol^{-1}. These results from the end-point calculations are also in similar agreement with experiments. While 3D-RISM in combination with QHIC or BQH lead to reasonable correlations with the PMF results and experiments, the calculated absolute binding free energies are underestimated

†Electronic supplementary information (ESI) available. See DOI: 10.1039/d1cp05075c

nanjie.deng@gmail.com .

Conflicts of interest

There are no conflicts to declare.

by ~5 kcal mol^{-1}. While the binding is accompanied by a significant reduction in the ligand translational/rotational entropy, the change in the torsional entropy in these host–guest systems is slightly positive. Compared with BQH, QHIC underestimates the reduction of configurational entropy because of the non-Gaussian probability distributions in the ligand rotation and a small number of torsions. The study highlights the crucial role of configurational entropy in determining binding and demonstrates the potential of using the new end-point method to provide insights in more complex protein–ligand systems.

Introduction

Molecular recognition plays vital roles in many biological processes such as enzyme catalysis, signal transduction, immune response, and gene regulation. Modeling molecular recognition at an atomistic detail can provide crucial insights for understanding the molecular mechanism underlying these biological phenomena. While many methods are available for computing the absolute and relative binding free energies,[1–8] our understanding of the molecular recognition will remain incomplete without a quantitative knowledge of the balance of various thermodynamic forces underlying binding. For example, DNA groove binders typically exhibit large entropy gain, whereas DNA intercalators generally show entropy loss.[9] What are the roles of solvent reorganization entropy and solute configurational entropy in these DNA-ligand systems? In another case, HIV-1 antivirals Darunavir (DRV) and Atazanavir (ATV) bind at the same pocket in the protease; however, while the former is driven by enthalpy, the latter is driven entirely by entropy.[10] The physical reasons for these fascinating behaviors have rarely been understood. Reliable computational tools to address such mechanistic questions will help inform ligand design beyond the estimation of binding affinity.

The receptor–ligand binding free energy can be computed using pathway methods such as the widely used alchemical pathway double decoupling method (DDM)[11–13] and physical pathway potential of mean force PMF[14] methods, and the more recently developed implicit ligand theory[15–17] and the alchemical transfer (ATM) method.[18] Alternatively, end-point methods such as MM-PB(GB)/SA,[19] LIE (Linear Interaction Energy),[20] and M2 (Mining Minima)[21] which need only consider the two end macrostates, are also popular in binding free energy calculations. While end-point methods generates a binding free energy estimate from the difference between two large numbers, which can lead to larger statistical error, such methods often allow for the binding free energy to be decomposed into physically intuitive components such as the direct intermolecular interaction, the solvation energy and entropy, and the solute configurational entropy, thus providing more direct insights into the thermodynamic driving forces of binding.[22] In contrast, in pathway based binding free energy methods, except for the state functions like the total entropy and enthalpy, the free energy components are dependent on the pathway. Furthermore, in terms of sampling, an end-point binding free energy calculation is not hampered by some of the sampling bottlenecks encountered in the pathway methods, such as the water occlusion problem in the ligand extraction simulations in PMF.

The successful estimation of the binding free energy using end-point methods requires accurate calculations of both the solvation free energy and the configurational entropy.[23–27] While MM-PB(GB)/SA is the most widely used end-point method,[22,27–30] its utility in dissecting the free energy contributions has been limited by the inadequate treatment of configurational entropy using approximate methods such as Normal Mode analysis (NMA). To provide a robust tool for analyze the thermodynamic determinants of binding, here we develop a new end-point method, called EE-BQH (Effective Energy-Boltzmann-Quasiharmonic), by incorporating an improved configurational entropy method Boltzmann-Quasiharmonic (BQH)[31] with different treatments of solvation free energy such as the continuum solvent PBSA[32] and the integral equation based 3D-RISM[33] to more rigorously calculate the absolute binding free energy. While BQH has been validated against theoretical configurational entropy computed using the exact Clausius expression,[31,34] it has yet to be used to compute binding free energy, which is one of the goals of the present study.

For realistic molecular systems, the direct calculation of the configurational entropy contribution ΔS^{config} using the full probability distribution function (PDF) $p(q_1, \ldots, q_N)$ is computationally intractable and many approximate methods have been developed.[24,25,35–40] In the quasiharmonic (QH) analysis,[35,37,41–43] the PDF is fitted to multivariate Gaussian which leads to overestimation of the calculated entropy when the underlying PDF is non-Gaussian.[36,37] To account for the non-Gaussian behavior, the leading term in the Boltzmann-Quasiharmonic method (BQH)[31,34,44] is the uncorrelated Gibbs entropy $-k_{\text{B}}\sum_i \int p(q_i)\ln p(q_i)\mathrm{d}q_i$, with the marginal PDF $p(q_i)$ estimated from the simulation data, which do not have to assume Gaussian. In both BQH and QH, the correlation among different degrees of freedom is estimated from the determinant of the normalized covariance matrix $C_{ij} = \sigma_{ij}/\left(\sigma_{ii}\sigma_{jj}\right)^{1/2}$ which can only capture linear pairwise correlations. While more rigorous treatments of the correlation are available in the mutual information expansion approaches MIE[24] and MIST,[45] these methods are computationally more demanding, as they require many two- and higher dimensional PDFs which can be difficult to converge. In contrast, BQH is computationally more feasible for realistic systems since the evaluation of covariance matrix does not involve two- or higher order PDFs. BQH has been tested in several model systems against the theoretically exact entropies from the Clausius equality and appears to strike a good balance between accuracy and computational complexity.[31,34] In this work we apply BQH in end-point absolute binding free energy calculations and compare its performance with other treatments of configurational entropy such as QHIC (Quasiharmonic model in Internal Coordinates), QHCC (Quasiharmonic model in Cartesian Coordinates), and NMA. The results are compared against those from the more accurate potential mean force (PMF) method[14,46,47] using the same force field to provide more meaningful tests of the different approaches.

The calculation of solvation free energy is another essential component in end-point binding free energy approaches. While PB and GB implicit solvent models are most frequently used in end-point binding free energy calculations, these methods treat the solvent as a featureless continuum dielectric medium and are known to introduce errors when the molecular nature of solvent molecules are important for binding.[27,29,30,48] On the other hand, the three-dimensional extension of the reference interaction site model (3D-RISM)[33,49,50] based on

the Ornstein–Zernike integral equation theory[51] can capture the inhomogeneous distribution and correlation of the molecular solvent.[52,53] In this study we compare the performance in end-point absolute binding free energies calculations of PBSA and 3D-RISM (the latter with linear partial molar volume corrections[54,55]) in combination with the configurational entropy approaches.

Because of their small sizes and simpler intermolecular energy landscapes compared to the more complex protein–ligand systems, host–guest systems allow for more complete sampling of the phase space and serve as useful model systems for validating and comparing binding free energy models.[21,36,56–59] Using host–guest systems as examples, Chang and Gilson *et al.* in an early study elucidated the roles of configurational entropy, preorganization, and induced fit effects in binding and observed strong compensation between the configurational entropy and effective energy.[21] Recently Chang's group reported the binding thermodynamics and kinetics of host–guest systems obtained using direct sampling of the association/dissociation event in microseconds MD and end-point calculations in explicit solvent.[60] Henchman and coworkers studied the binding thermodynamics of the CB8 host–guest systems as part of the SAMPL8 challenge using an end-point method in which the total entropy is calculated in a hierarchical manner.[61]

In the present study, we validate the EE-BQH end-point method by applying it to compute the absolute binding free energies for the host–guest systems from the recent SAMPL8 (Statistical Assessment of the Modeling of Proteins and Ligands) GDCC Host–Guest Challenge set[62] and compare the results with other treatments of configurational entropy such as QHIC, QHCC, and NMA. We found that the results obtained using BQH and QHIC show the best correlations with the absolute binding free energies computed using the more accurate pathway-based potential of mean force method (PMF),[8,46,47,63–65] with R^2 of 0.69–0.77 and averaged unsigned errors of 1.53–1.88 kcal mol^{-1}. These results from the end-point calculations are also in similar agreement with the experimental measurements. Interestingly, while the binding is opposed by a large reduction in ligand external entropy, these host–guest systems exhibit a small increase in the torsional entropy upon binding. By dissecting the binding free energy into contributions from solute configurational entropy and effective potential energy, the EE-BQH model provides thermodynamic insights for analyzing the driving forces of binding not easily available from pathway methods.

Methods and materials

The end-point formula for absolute binding free energy[11] can be derived from the solute chemical potential for dilute solutions[66]

$$\mu_i = -kT\ln\left(\frac{8\pi^2}{C^\circ\prod_{j=1}^{M}\Lambda_j^3}\int \mathrm{d}q\, e^{-\phi(q)/kT}\right) + kT\ln\frac{C_i}{C^\circ} \quad (1)$$

where Λ_j is the de Broglie thermal wavelength of the atom j in the solute molecule i, $\phi(q)$ is the effective potential energy of a solute molecule in solution as a function of the $3M-6$ internal coordinates q of the solute (M is the number of atoms in a solute molecule).

Eqn (1) can be derived in different ways, such as using the grand canonical ensemble,[67] but the most straightforward way involves using the Widom insertion formula[68] as shown below:

Consider a binary solution containing N_1 solvent molecules 1 and N_2 solute molecules 2. In the canonical ensemble, the Helmholtz free energy $A = -kT\ln Q_{N1,N2}$ with $Q_{N1,N2} = \frac{1}{N_1!\Lambda_1^{3N_1}}\frac{1}{N_2!\Lambda_2^{3N_2}}\int e^{-\beta U\left(r^{N_1}, R^{N_2}\right)}\mathrm{d}r^{N_1}\mathrm{d}R^{N_2}$. Standard manipulation yields

$$\begin{aligned}\mu_2 = \left(\frac{\partial A}{\partial N_2}\right)_{N_1,V,T} &= -kT\ln\frac{Q_{N_1,N_2}}{Q_{N_1,N_2-1}} \\ &= kT\ln N_2\Lambda_2^3 - kT\ln\frac{Z_{N_1,N_2}}{Z_{N_1,N_2-1}}\end{aligned} \tag{2}$$

where r^{N_1} and R^{N_2} denote the solvent and solute coordinates, respectively; $Z_{N_1,N_2} = \int \mathrm{e}^{-\beta U\left(r^{N_1}, R^{N_2}\right)}\mathrm{d}r^{N_1}\mathrm{d}R^{N_2}$, and $Z_{N_1,N_2-1} = \int \mathrm{e}^{-\beta U\left(r^{N_1}, R^{N_2-1}\right)}\mathrm{d}r^{N_1}\mathrm{d}R^{N_2-1}$. Applying the Widom insertion formula, the system potential energy $U\left(r^{N_1}, R^{N_2}\right)$ can be written as

$$U\left(r^{N_1}, R^{N_2}\right) = U\left(r^{N_1}, R^{N_2-1}\right) + u\left(R_\alpha, r^{N_1}, R^{N_2-1}\right) \tag{3}$$

where R_α denotes the coordinates of an arbitrarily chosen solute molecule α and $u\left(R_\alpha, r^{N_1}, R^{N_2-1}\right)$ is the perturbation energy of inserting the solute α into the solution containing N_1 solvent molecules 1 and $N_2 - 1$ solute molecules

$$u\left(R_\alpha, r^{N_1}, R^{N_2-1}\right) = U\left(r^{N_1}, R^{N_2}\right) - U\left(r^{N_1}, R^{N_2-1}\right)$$

Inserting eqn (4) into the expression of Z_{N_1,N_2} and integrating out the six external degrees of freedom of the distinguished molecule α yields

$$Z_{N_1,N_2} = 8\pi^2 V\int \mathrm{d}q_\alpha \int \mathrm{e}^{-\beta U\left(r^{N_1}, R^{N_2-1}\right)}\mathrm{e}^{-\beta u\left(q_\alpha, r^{N_1}, R^{N_2-1}\right)}\mathrm{d}r^{N_1}\mathrm{d}R^{N_2-1} \tag{4}$$

It follows from eqn (4) that

$$\begin{aligned}&-kT\ln\frac{Z_{N_1,N_2}}{Z_{N_1,N_2-1}} \\ &\quad = -kT\ln 8\pi^2 V\frac{\int \mathrm{d}q_\alpha\int \mathrm{e}^{-\beta U\left(r^{N_1}, R^{N_2-1}\right)}\mathrm{e}^{-\beta u\left(q_\alpha, r^{N_1}, R^{N_2-1}\right)}\mathrm{d}r^{N_1}\mathrm{d}R^{N_2-1}}{\int \mathrm{e}^{-\beta U\left(r^{N_1}, R^{N_2-1}\right)}\mathrm{d}r^{N_1}\mathrm{d}R^{N_2-1}}\end{aligned} \tag{5}$$

For an infinitely dilute solution, $N_2 - 1 \rightarrow 0$, and eqn (5) becomes

$$-kT\ln\frac{Z_{N_1,N_2}}{Z_{N_1,N_2-1}} = -kT\ln 8\pi^2 V \int \mathrm{d}q_\alpha \frac{\int \mathrm{e}^{-\beta U\left(r^{N_1}\right)}\mathrm{e}^{-\beta u\left(q_\alpha, r^{N_1}\right)}\mathrm{d}r^{N_1}}{\int \mathrm{e}^{-\beta U\left(r^{N_1}\right)}\mathrm{d}r^{N_1}} \tag{6}$$

Inserting eqn (6) into eqn (2) yields

$$\mu_2 = -kT\ln\frac{8\pi^2}{\Lambda_2^3 C} - kT\ln\frac{\int \mathrm{e}^{-\beta U\left(r^{N_1}\right)}\mathrm{e}^{-\beta u\left(q_\alpha, r^{N_1}\right)}\mathrm{d}r^{N_1}}{\int \mathrm{e}^{-\beta U\left(r^{N_1}\right)}\mathrm{d}r^{N_1}} \tag{7}$$

$\left(C = \frac{N_2}{V}\right)$. Define the effective energy $\phi(q)$ as

$$\mathrm{e}^{-\phi(q_\alpha)/kT} = \frac{\int \mathrm{e}^{-\beta U\left(r^{N_1}\right)}\mathrm{e}^{-\beta u\left(q_\alpha, r^{N_1}\right)}\mathrm{d}r^{N_1}}{\int \mathrm{e}^{-\beta U\left(r^{N_1}\right)}\mathrm{d}r^{N_1}} \tag{8}$$

It follows from eqn (7) and (8) that

$$\mu_2 = -kT\ln\frac{8\pi^2}{\Lambda_2^3 C} - KT\ln\int \mathrm{d}q_\alpha e^{-\phi(q_\alpha)/kT},$$

which is eqn (1).

For the binding reaction A + B $\rightleftharpoons$ AB, inserting eqn (1) into the equilibrium condition $\mu_A + \mu_B = \mu_{AB}$ yields

$$\begin{aligned}\Delta G^\circ_{\mathrm{bind}} \equiv & -kT\ln\left(\frac{C_{\mathrm{AB}}C^\circ}{C_{\mathrm{A}}C_{\mathrm{B}}}\right) \\ = & -kT\ln\frac{C^\circ}{8\pi^2} - kT\ln\int_{V_{\mathrm{site}}}\mathrm{d}q_{\mathrm{AB}}\mathrm{e}^{-\phi(q_{\mathrm{AB}})/kT} \\ & + kT\ln\int \mathrm{d}q_{\mathrm{A}}\mathrm{e}^{-\phi(q_{\mathrm{A}})/kT} + kT\ln\int \mathrm{d}q_{\mathrm{B}}\mathrm{e}^{-\phi(q_{\mathrm{B}})/kT}\end{aligned} \tag{9}$$

Note that the momentum energy terms Λ_j^3 in μ_{AB}, μ_A and μ_B cancel. The phase space integration for the complex AB is limited to the sub-volume V_{site} which defines the complexed state in the six external dimensions of B relative to A.[46]

Since

$$-kT\ln\int \mathrm{d}q\mathrm{e}^{-\phi(q)/kT} = \int \mathrm{d}q p(q)\phi(q) + kT\int \mathrm{d}q p(q)\ln p(q) \tag{10}$$

where $p(q)$ is the probability density of finding solute i in configuration q in solution

$$p(q) = \frac{\mathrm{e}^{-\phi(q)/kT}}{\int \mathrm{d}q\mathrm{e}^{-\phi(q)/kT}},$$

it follows from eqn (10) and (9) that

$$\begin{aligned}\Delta G^{\circ}_{\mathrm{bind}} &= \Delta\langle\phi\rangle - T\Delta S^{\circ}_{\mathrm{config}} \\ &= (\langle\phi\rangle_{\mathrm{AB}} - \langle\phi\rangle_{\mathrm{A}} - \langle\phi\rangle_{\mathrm{B}}) \\ &\quad - T\left(S_{\mathrm{config,AB}} - S_{\mathrm{config,\,A}} - S_{\mathrm{config,B}} + k_{\mathrm{B}}\ln\frac{C^{\circ}}{8\pi^2}\right)\end{aligned} \tag{11}$$

where $\langle\phi\rangle_X = \int \mathrm{d}q\phi(q)p(q)$, $S_{\mathrm{config,X}} = -k\int \mathrm{d}q p(q)\ln p(q)$ and the total configurational entropy of binding is given by

$$\Delta S^{\circ}_{\mathrm{config}} = S_{\mathrm{config,AB}} - S_{\mathrm{config,A}} - S_{\mathrm{config,B}} + k_{\mathrm{B}}\ln\frac{C^{\circ}}{8\pi^2} \tag{11a}$$

Eqn (11) is the end-point binding free energy formula. Note that the above derivation does not assume the pairwise additivity in the interaction potential and is therefore applicable for both classical additive force fields and polarizable force fields.

In the present study the potential used is pairwise additive, in which case the effective energy $\phi(q)$ can be written as the sum of intramolecular energy $u_{\mathrm{intra}}(q)$ and solvation free energy $w(q)$, *i.e.,*

$$\phi(q) = u_{\mathrm{intra}}(q) + w(q) \tag{11b}$$

where the solvent potential of mean force $w(q)$ is

$$\mathrm{e}^{-w(q)/kT} = \frac{1}{Z_{N,0}}\int \mathrm{d}r^{N_1}\mathrm{e}^{-\left[u_{uv}\left(q,r^{N_1}\right)+U_{vv}\left(r^{N_1}\right)\right]/kT} \tag{12}$$

Here $u_{uv}\left(q,r^{N_1}\right)$ is the interaction between one solute with the N_1 solvent molecules, and $U_{\mathrm{vv}}\left(r^{N_1}\right)$ denotes the solvent–solvent interactions. $Z_{N,0} = \int \mathrm{d}r^{N_1}e^{-U_{\mathrm{vv}}\left(r^{N_1}\right)/kT}$ is the configurational integral of the pure solvent.

To obtain the effective energy $\langle\phi_i\rangle$ requires calculating the solute internal energy $\langle u_{\mathrm{intra}}(q)\rangle$ and solvation free energy $\langle w(q)\rangle$. The latter can be computed for the solute in the fixed configuration q as

$$w(q) = \left[\int \mathrm{d}r^{N1} P\left(r^{N1} \mid q\right)\left(u_{\mathrm{uv}}\left(q, r^{N1}\right) + U_{\mathrm{vv}}\left(r^{N1}\right)\right) - \int \mathrm{d}r^{N1} P_0\left(r^{N1}\right) U_{\mathrm{vv}}\left(r^{N1}\right)\right] - T\left[-k\int \mathrm{d}r^{N1} P\left(r^{N1}|q\right)\ln P\left(r^{N1}|q\right) + k\int \mathrm{d}r^{N1} P_0\left(r^{N1}\right)\ln P_0\left(r^{N1}\right)\right] \tag{13}$$

where $P\left(r^{N1}|q\right) = P\left(q, r^{N1}\right)/P(q)$ is the conditional probability density of finding solvent in the r^{N1} configuration given that the solute configuration is q. $P_0\left(r^{N1}\right)$ is the probability density in pure solvent. The first and second brackets in the eqn (13) corresponds to the solvation enthalpy $\Delta H_{\mathrm{solv}}(q)$ and solvation entropy $-T\Delta S_{\mathrm{solv}}(q)$, respectively. While several methods exist to compute separately the solvation entropy $\Delta H_{\mathrm{solv}}(q)$ and solvation entropy $-T\Delta S_{\mathrm{solv}}(q)$,[60,61,69] since the main goal of this study is to improve end-point calculations of absolute binding free energy by using better configurational entropy treatments, here we apply two approximate solvation free energy methods, the continuum solvent based PBSA[29,30] and the integral equation based 3D-RISM,[70] on the configurations sampled from explicit solvent MD to estimate the solvation free energy $\langle w(q)\rangle$. The resulting $\langle w(q)\rangle$ are used in eqn (11b) to estimate the effective energy $\langle\phi\rangle$ and the absolute binding free energy.

Quasiharmonic model in internal coordinates (QHIC)

In the Quasiharmonic approximation,[35] the multidimensional probability density of a molecule with n internal degrees of freedom is assumed to be normalized multivariate Gaussian

$$p(q) = \frac{1}{(2\pi)^{n/2}\left|\sigma\right|^{1/2}} \mathrm{e}^{-\frac{1}{2}\left[(q-\langle q\rangle)^{\mathrm{T}}\sigma^{-1}(q-\langle q\rangle)\right]} \tag{14}$$

where σ is the covariance matrix, $\sigma_{ij} = \left\langle\left(q_i - \langle q_i\rangle\right)\left(q_j - \langle q_j\rangle\right)\right\rangle$, of the internal coordinates (including for the complex the six ligand external degrees of freedom X_{L}, Y_{L}, Z_{L}, θ, φ, and ψ relative to the receptor). $|\sigma|$ is the determinant of σ. Substituting the PDF of eqn (14) into the Gibbs entropy expression yields the configurational entropy of the molecule

$$S^{\mathrm{QHIC}} = -k\int \mathrm{d}q p(q)\ln p(q) = \frac{1}{2}nk + \frac{1}{2}k\ln\left[(2\pi)^n\left|\sigma\right|\right] \tag{15}$$

And the configurational entropy change due to binding is given by

$$\Delta S^{\circ}_{\mathrm{config}} = -k\ln 8\pi^2 V^{\circ} + S^{\mathrm{QHIC}}_{\mathrm{AB}} - S^{\mathrm{QHIC}}_{\mathrm{A}} - S^{\mathrm{QHIC}}_{\mathrm{B}} \tag{16}$$

Note that while S_{AB}^{QHIC} includes both the six guest external degrees of freedom and all the torsions in both host and guest molecules, S_{A}^{QHIC} and S_{B}^{QHIC} only include their torsional degrees of freedom. The determinant $|\sigma|$ in eqn (15) is calculated using the MATLAB library.

Boltzmann Quasiharmonic model (BQH)

It is well known that fluctuations in certain degrees of freedom such as a torsion with several rotamer states can be strongly non-Gaussian and the application of QHIC approximation eqn (15) in such cases will overestimate the configurational entropy.[36,37] The Boltzmann Quasiharmonic model (BQH) was proposed to better account for the non-Gaussian behavior in estimating entropy.[31] Note that the determinant $|\sigma|$ in eqn (15) can be factored into the product of the diagonal elements of σ and the determinant of the normalized fluctuation covariance matrix C

$$\left|\sigma\right| = \prod_{i}^{n} \sigma_{ii} \left|C\right| \tag{17}$$

where $C_{ij} = \sigma_{ij}/\left(\sigma_{ii}\sigma_{jj}\right)^{1/2}$. Therefore, eqn (15) can be written as

$$S^{QHIC} = \left\{\frac{1}{2}nk + \frac{1}{2}k\ln\left[(2\pi)^{n}\prod_{i}^{n}\sigma_{ii}\right]\right\} + \frac{1}{2}k\ln|C| \tag{18}$$

It is easy to see that first term in the right-hand side of eqn (18) is the uncorrelated first order Boltzmann entropy, *i.e.* $\frac{1}{2}nk + \frac{1}{2}k\ln\left[(2\pi)^{n}\prod_{i}^{n}\sigma_{ii}\right] = -k_{B}\sum_{i}^{n}\int p(q_i)\ln p(q_i)\mathrm{d}q_i$, when the distribution along each q_i is Gaussian, *i.e.* $p(q_i) = \frac{1}{\sigma_{ii}\sqrt{2\pi}}\mathrm{e}^{-\frac{(q_i - q_i)^2}{2\sigma_{ii}^2}}$. When the distributions $p(q_i)$ are non-Gaussian, a more accurate treatment is to replace the first term in eqn (18) with the uncorrelated first order Boltzmann entropy $-k_{B}\sum_{i}^{n}\int p(q_i)\ln p(q_i)\mathrm{d}q_i$, rather than estimating it from the variances σ_{ii} as in the QHIC approximation. Thus, in BQH, the configurational entropy of a molecule is given by

$$S^{BQH} = -k_{B}\sum_{i}^{n}\int p(q_i)\ln p(q_i)\mathrm{d}q_i + \frac{1}{2}k\ln|C| \tag{19}$$

Note that the treatment of the correlation between degrees of freedom are the same as in QHIC and can only account for linear pairwise correlations.

The configurational entropy of binding from BQH is calculated by substituting eqn (19) into eqn (11a)

$$\Delta S^{\circ}_{config} = -k\ln 8\pi^{2}V^{\circ} + S_{AB}^{BQH} - S_{A}^{BQH} - S_{B}^{BQH} \tag{19a}$$

$\Delta S^{\circ}_{\text{config}}$ can be written as the sum of the uncorrelated Boltzmann entropy change ΔS_1 and the second order correlated entropy contribution ΔS_2

$$\Delta S^{\circ}_{\text{config}} \approx \Delta S_1 + \Delta S_2 \tag{20}$$

where

$$\Delta S_1 = \left(S_{1,\text{AB}} + k_{\text{B}}\ln\frac{C^{\circ}}{8\pi^2}\right) - S_{1,\text{A}} - S_{1,\text{B}} \tag{21}$$

and

$$\Delta S_2 = S_{2,\text{AB}} - S_{2,\text{A}} - S_{2,\text{B}} \tag{22}$$

In this study, we focus on the soft degrees of freedom that contribute most to ΔS_{config}, which includes torsion angles and the ligand translational/rotational degrees of freedom, since the contributions from the bond stretching and angle bending are typically much smaller than those from the soft degrees of freedom[31,71] and can be safely ignored. Thus

$$\begin{aligned}\Delta S_1 &= \left(S_{1,\text{AB}} + k_{\text{B}}\ln\frac{C^{\circ}}{8\pi^2}\right) - S_{1,\text{A}} - S_{1,\text{B}} \\ &= \Delta S_1^{\text{trans/rot}} + \Delta S_1^{\text{torsion}}\end{aligned} \tag{23}$$

Here $\Delta S_1^{\text{trans/rot}}$ is the change in the uncorrelated translational/rotational entropy from the complexation and is written in discretized form as

$$\begin{aligned}\Delta S_1^{\text{trans/rot}} &= S_{1,\text{AB}}^{\text{trans/rot}} - k_{\text{B}}\ln\left(\frac{8\pi^2 V^{\circ}}{\prod_q \Delta q}\right), \\ S_{1,\text{AB}}^{\text{trans/rot}} &= -k\sum_q\sum_i^{N_{\text{bin}}} p_i^q \ln p_i^q \quad (q = X_{\text{L}}, Y_{\text{L}}, Z_{\text{L}}, \theta, \varphi, \psi)\end{aligned} \tag{24}$$

where $S_{1,\text{AB}}^{\text{trans/rot}}$ is the ligand translational and rotational entropies within the complex, as described by the six external degrees of freedom of the bound ligand, *i.e.*, the ligand center of mass $(X_{\text{L}}, Y_{\text{L}}, Z_{\text{L}})$ defined in the reference frame of the receptor, and the three Euler angles (θ, φ, ψ) defined by choosing two receptor atoms and three ligand atoms as described previously.[46,47] $-k_{\text{B}}\ln\left(\frac{8\pi^2 V^{\circ}}{\prod_q \Delta q}\right)$ in eqn (18) is the translational/rotational entropy of the reference state for a free ligand in a solution at the standard concentration $C^{\circ} = \frac{1}{V^{\circ}}$, which has a uniform distribution over the N_{bin} bins along each of the six external degrees of freedom. Here Δq is the bin width of the corresponding degree of freedom q. In this work, N_{bin} is set to be 100 for each degree of freedom.

For the uncorrelated torsional entropy change

$$\Delta S_1^{\text{torsion}} = S_{1,\text{AB}}^{\text{torsion}} - S_{1,\text{A}}^{\text{torsion}} - S_{1,\text{B}}^{\text{torsion}}$$
$$S_{1,\text{X}}^{\text{torsion}} = -k \sum_{\{\text{torsions}\}} \sum_{i}^{N_{\text{bin}}} p_i^q \ln p_i^q \quad (25)$$

where $S_{1,\text{X}}^{\text{torsion}}$ is the torsional entropy of each species in solution.

For the second order correlation contribution to $S^{\circ}_{\text{config}}$, S_2 is calculated using eqn (19), *i.e.*

$$S_{2,\text{X}} = \frac{1}{2} k \ln |C|,$$

where $C_{ij} = \sigma_{ij}/\left(\sigma_{ii}\sigma_{jj}\right)^{1/2}$ as defined in eqn (17). Note that while $S_{2,\text{AB}}$ includes the correlations among the ligand translations and rotations within the bound complex and all the torsions in the two molecules, $S_{2,\text{A}}$ and $S_{2,\text{B}}$ only include the correlations among the torsional degrees of freedom within each of the host and guest molecules. To avoid the problem of periodicity in the torsion angles in calculating the covariance matrix σ_{ij}, a torsion angles φ is represented as $e^{i\varphi}$,[31] and the corresponding covariance matrix $\sigma_{mn} = \left\langle \left(e^{i\varphi m} - \left\langle e^{i\varphi m} \right\rangle\right)\left(e^{-i\varphi n} - \left\langle e^{-i\varphi n} \right\rangle\right) \right\rangle$ is guaranteed to have real determinant because it is Hermitian.

Quasiharmonic model in Cartesian coordinates (QHCC) and Normal Mode analysis (NMA)

In both QHCC and NMA, the configurational entropy change due to binding is approximated as[42,43,72]

$$\Delta S^{\circ}_{\text{config}} = \Delta S^{\text{trans}} + \Delta S^{\text{rot}} + \Delta S^{\text{vibr}} \quad (26)$$

where ΔS^{trans} and ΔS^{rot} are estimated using the ideal gas entropy (Sackur–Tetrode) at standard concentration C° and the rigid-rotor expressions,[73] respectively. In both methods, the S^{vibr} is calculated using the entropy expression for harmonic oscillators

$$\Delta S^{\text{vibr}} = \sum_{i=1}^{3N-6} k \left[\frac{h\upsilon_i/kT}{e^{-h\upsilon_i/kT} - 1} - \ln\left(1 - e^{-\frac{h\upsilon_i}{kT}}\right) \right] \quad (27)$$

where υ_i is the frequency of the i-th harmonic oscillator. In NMA, υ_i is calculated from the eigenvalue of the mass-normalized second derivative matrix for energy minimized structures. In QHCC, after removing the overall translation and rotation of the solute, the eigenvalues of the mass-normalized covariance matrix are solved to obtain υ_i. Both QHCC and NMA entropies are calculated using the implementation in the Amber package.[74]

MD Simulation of the host–guest systems

In this work, the MD free energy simulations were performed using the PMEMD program in the AMBER20 package.[74] The atomic coordinates of the ten bound complexes of the

two octa acids hosts, tetramethyl octa acid (TEMOA) and tetraethyl octa acid (TEETOA), and five guest molecules were obtained by manually docking the guests to each host using Maestro (Schrodinger Inc.), see Fig. 1. The TEETOA host differ from TEMOA by the replacement of a methyl at the cavity rim with an ethyl. The Amber GAFF2 parameters set[75] and the AM1-BCC charge model[76] are used to describe the solute. The dimension of the solvent box is set up to ensure that the distance between solute atoms from nearest walls of the box is at least 12 Å and Na^+ ions are added to the solvent box to maintain charge neutrality. The system equilibration involved the following series of minimization and molecular dynamics steps: the structure was first minimized and heated from 0 to 300 K in 200 ps with harmonic restraints on the heavy atoms and a force constant of 10 kcal $Å^{-2}$ mol under constant volume conditions, followed by four MD simulations with gradually reduced restraints: (1) 3 ns with $k = 5$ kcal $Å^{-2}$ mol, (2) 3 ns with $k = 1$ kcal $Å^{-2}$ mol, (3) 3 ns with $k = 0.25$ kcal $Å^{-2}$ mol^{-1}, and (4) 3 ns unrestrained MD under constant pressure conditions. Finally, a production MD of 30 ns was performed using the *NPT* ensemble and Langevin thermostat with isotropic position scaling for constant pressure. The electrostatic interactions were computed using the particle–mesh Ewald (PME)[77] method with a real space cutoff of 9.0 Å and a grid spacing of 1.0 Å. MD simulations were performed in the *NPT* ensemble with a time step of 2 fs. For configurational entropy calculation, MD trajectories are saved every 0.1 ps, with a total of 300 000 trajectory frames used for one entropy calculation.

Solvation free energy calculations

The solvation free energy $w(q)$ in eqn (12) is calculated using two approximate methods, the continuum solvent based PBSA[29,30] and the integral equation based 3D-RISM,[70] on the MD structures sampled from three independent simulations for the three species (complex, host and guest) in explicit solvent. In 3D-RISM, the solvation free energy is calculated using the rism3d.snglpnt program available from the AMBER20 package[74] on the solute structures extracted from the explicit solvent MD simulations of the three species in binding; the KH closure and the linear partial molar volume correction[54,55] are used in the rism3d.snglpnt calculations. We found that the use of the partial molar volume correction[54,55] leads to substantially improved absolute binding free energies predictions (compare Fig. 3 with Fig. S2, ESI†).

The PBSA solvation free energy calculations were performed using the MMPBSA.py[32] tool from the AMBER20[74] package. The solvation free energies of the complex, host and guest are computed using 600 trajectory frames (*i.e.*, every 50 ps) of solute structures extracted from the corresponding 30 ns MD trajectories in explicit solvent. The ionic concentration was set to 0.1 M for all calculations. Error bars were generated by comparing binding energy components of the first and second half of the MD simulations.

Results

Absolute binding free energies from the different end-point methods

We performed end-point calculations using different combinations of configurational entropy and solvation free energy models to estimate the absolute binding free energy

of the ten SAMPL8 host–guest complexes. Since comparing the calculated binding free energies with the experiments can be complicated by the errors from the energy functions, to minimize such complications in evaluating a new method we first compare the end-point results with those obtained using the more well established PMF method in explicit solvent with the same force field, see Table 1 and Fig. 2. The latter were found in excellent agreement with the experimental measurements of the absolute binding free energies.[65] The comparisons between the end-point calculations with the experimental results are given in the Table S1 and Fig. S1 (ESI†).

Table 1 summarizes the overall results compared with the theoretical benchmark of the absolute binding free energies obtained from the PMF calculations in explicit solvent.[65] Fig. 2 provides a graphical view of the overall performance of different combinations. QHIC/PBSA and BQH/PBSA are the only endpoint methods that produced absolute binding free energy estimates with an averaged error smaller than 2.0 kcal mol^{-1} compared with the PMF results. The two methods also resulted in good correlations with the PMF-calculated absolute binding free energies in terms of rank-ordering, with the R^2 of 0.69 for BQH/PBSA and 0.77 for QHIC/PBSA. Both methods correctly predicted that every guest molecule binds more strongly with TEMOA host than TEETOA host, and that TEMOA-G2 shows the strongest binding affinity. Therefore, despite the use of implicit solvent model for solvation free energy, the results from the end-point methods QHIC/PBSA and BQH/PBSA are in good agreement with those obtained using the pathway based PMF method in explicit solvent simulations. Besides QHIC/PBSA and BQH/PBSA, the results from QHCC/PBSA also shows a reasonable correlation with the PMF calculations, with $R^2 = 0.48$ and averaged error 2.65 kcal mol^{-1}.

While the combinations QHIC/3D-RISM and BQH/3D-RISM produced reasonably good correlations with the PMF results with R^2 of 0.63 and 0.48, respectively, the averaged unsigned errors in the absolute binding free energies obtained by using the 3D-RISM solvation model are significantly larger than those obtained from QHIC/PBSA and BQH/PBSA calculations using the continuum solvent PBSA (Table 1 and Fig. 2). Table 2 reveals that on average the combination of 3D-RISM with the different configurational entropy methods systematically underestimated the absolute binding free energies by ≥ 5.1 kcal mol^{-1}. Unsurprisingly, the calculations using the more approximate NMA for configurational entropy in combination with either of the solvation methods resulted in poor correlations with the theoretical benchmark of the PMF calculations and very large unsigned errors in the calculated absolute binding free energies (Table 2 and Fig. 3). Overall, using the PMF calculated absolute binding free energies as the theoretical benchmark shows that for the host–guest systems tested here, both BQH and QHIC in combination with PBSA lead to more accurate absolute binding free energy estimates than other approaches tested.

Comparing these results with experiments lead to essentially the same conclusion that BQH and QHIC in combination with PBSA yield reasonably good absolute binding free energy estimates: see Table S1 (ESI†). Both the R^2 and the averaged unsigned errors obtained using the experimental results as the reference are comparable to those obtained using the PMF results as the reference. We also note that the predictions from BQH/PBSA and QHIC/PBSA also compare favorably to the results reported in the SAMPL8 challenge using other

binding free energy methods (Table S2, ESI†). Among the nine sets of results obtained using nonpolarizable force fields (AMOEBA uses a polarizable charge model), the BQH/PBSA and QHIC/PBSA are ranked at top 3 and top 2 places, respectively in terms of R^2.

Insights into the host–guest binding from the entropic components from BQH

We analyze the components of the solute configurational entropy from BQH to gain insights into the host–guest binding. First, we check the convergence of the BQH calculated configurational entropy change on binding. Fig. 4 shows the time behavior of the $T\Delta S^{\circ}_{\mathrm{config}}$ components for the G1 guest molecule in complex with both TEETOA and TEMOA hosts. While the uncorrelated torsional entropies show small drifts of the order of ~0.3 kcal mol^{-1}, the translational/rotational entropy and the second order correlation contributions appear to be more stable starting from t = 15 ns.

Table 3 lists the configurational entropy contributions calculated from BQH for the host–guest complexes. The uncorrelated ligand translation/rotation entropy accounts for the bulk of the configurational entropy loss due to binding. According to Table 3, the bulk of the total solute entropy change $\Delta S^{\circ}_{\mathrm{config}}$ comes from the 1st order uncorrelated entropy ΔS_1, while the correlation among the different degrees of freedom, *i.e.*, $\Delta S_2^{\mathrm{corr}}$ accounts for between 10% and 30% of the total configurational entropy loss upon binding. There is no clear-cut relationship between the uncorrelated entropy ΔS_1 and the 2nd order correlation entropy $\Delta S_2^{\mathrm{corr}}$.

Some of the configurational entropy values in Table 3 can be intuitively related to the chemical structures. For example, the complexes TEETOA-G3 and TEMOA-G3 both exhibits relatively small configurational entropy losses, which could be explained by the fact that G3 has the smallest and more flexible ring comparing with the other guest molecules (Fig. 1). Interestingly, for both TEETOA and TEMOA hosts, the guest binding is accompanied by a small increase in the uncorrelated torsional entropy $\Delta S_1^{\mathrm{torsion}}$, which ranges from 0.31 to 1.67 kcal mol^{-1} (Table 3). Fig. 5 shows the histograms extracted from MD trajectories of a torsion angle in the unliganded host and in the complex, which illustrates the increase in the torsion entropy upon binding. In this example, the distribution of the torsion in the unliganded host is dominated by a peak centered at 100°, whereas in the complex the distribution is more even (Fig. 5).

Comparison of BQH *vs*. QHIC calculated entropies

As shown in the Methods, BQH and QHIC are two closely related methods. While the configurational entropy changes calculated by these two best performing methods are highly correlated with each other (Fig. 6), QHIC consistently predicts smaller entropy losses upon binding ($T\Delta S^{\mathrm{QHIC}}_{\mathrm{config}}$) than BQH does by up to 1.28 kcal mol^{-1} (Table 4). As will be shown below, the smaller entropy loss predicted by QHIC is the result of its overestimation of the configurational entropy due to the non-Gaussian nature of the PDF along the ligand rotational and a small number of torsional degrees of freedom (Fig. 7 and 8). Even before we delve into the PDFs, there is evidence suggesting that the BQH calculated entropy

change $T\Delta S^{\mathrm{BQH}}_{\mathrm{config}}$ is more correct. For example, when comparing the TEETOA-G3 and TEETOA-G1 complexes, the smallest and most flexible guest molecule G3 is correctly predicted to experience the smaller entropy loss upon binding by BQH (Table 4). According to the QHIC calculations, however, G3 shows larger entropy loss than G1, which is likely a result of the overestimation of the entropy as shown in detail below.

Since $\Delta S^{\circ}_{\mathrm{config}}$ is largely dominated by the ligand translation and rotation entropy contributions (Table 3), we first examine the marginal PDF along the three translational and three rotational degrees of freedom of the bound ligand to understand the physical reasons for the differences between the entropies calculated using QHIC and BQH. Fig. 7 shows the results for the guest molecules in the TEMOA-G3 and TEMOA-G4 complexes, which are representative for the ten complexes in Table 4, as these two complexes show the smallest and largest differences between the QHIC- and BQH-calculated entropies $\Delta S^{\mathrm{QHIC}}_{\mathrm{config}} - \Delta S^{\mathrm{BQH}}_{\mathrm{config}}$. As seen from Fig. 7, while the PDFs for the ligand translations inside the complex can be well described by single Gaussian, the ligand rotations can deviate sharply from unimodal Gaussian distribution, especially in the TEMOA-G4 complex. Consequently, while BQH and QHIC predict very similar translational entropy losses upon binding $\Delta S^{\mathrm{QHIC}}_{\mathrm{trans}}$ and $\Delta S^{\mathrm{BQH}}_{\mathrm{trans}}$ ($\Delta S_{\mathrm{trans}} = S_{\mathrm{trans}} - k\ln V^{\circ}$) that differ by <0.02 kcal mol^{-1}, QHIC overestimates the rotational entropy compared those from BQH, resulting in smaller rotational entropy losses given by $\Delta S^{\mathrm{QHIC}}_{\mathrm{rotation}}$ ($\Delta S^{\mathrm{QHIC}}_{\mathrm{rotation}} = \Delta S^{\mathrm{QHIC}}_{\mathrm{rotation}} - k\ln 8\pi^2$), see Fig. 7, right panel. For example, in the TEMOA-G3 complex, the ligand rotational PDFs contain multimodal peaks that are not well separated and can still be fitted approximately by a single Gaussian; in this case the overestimation in the calculated rotational entropies by using the QHIC will be relatively small, with $T\left(\Delta S^{\mathrm{QHIC}}_{\mathrm{rotation}} - \Delta S^{\mathrm{BQH}}_{\mathrm{rotation}}\right) = 0.195$ kcal mol^{-1}. In the TEMOA-G4 complex, however, the three ligand rotational PDF contain multiple well-separated peaks; using unimodal Gaussian fitting therefore causes a significantly larger overestimation in entropy,[37] in this case $T\left(\Delta S^{\mathrm{QHIC}}_{\mathrm{rotation}} - \Delta S^{\mathrm{BQH}}_{\mathrm{rotation}}\right) = 0.626$ kcal mol^{-1} (Fig. 7).

In addition to the three ligand rotational degrees of freedom, the non-Gaussian behavior in a small number of torsional degrees of freedom also contributes to the overestimated configurational entropy by QHIC, but to a smaller extent compared with the ligand rotations discussed above. This is because first, only a very small number of torsion angles in these host–guest systems exhibit non-Gaussian, multimodal distributions. Fig. 8 shows one of such torsions that exhibit non-Gaussian multimodal distributions in both complexed and unbound state. Second, for those few torsions that show multimodal distributions, the entropies for bound state and unbound state will both be overestimated by QHIC, and the net effect on binding entropy is small because of error cancellation. For example, the strongly non-Gaussian distribution in the torsion angle $\chi 1$ in Fig. 8 leads to just 0.295 kcal mol^{-1} for the difference between the QHIC- and BQH-calculated entropy changes $T\left(\Delta S^{\mathrm{QHIC}}_{\mathrm{config}} - \Delta S^{\mathrm{BQH}}_{\mathrm{config}}\right)$.

In summary, the smaller losses of entropy due to binding predicted by QHIC (Table 4) are attributable to the overestimation of the ligand rotational entropy and the torsional entropy by QHIC, which uses single Gaussian to approximate the underlying PDF that can be strongly non-Gaussian. These results show that a histogram approach like the one implemented in BQH is required to more accurately account for non-Gaussian behavior to compute the uncorrelated entropies.

Lastly, we note that although QHIC overestimated the configurational entropies compared with the more accurate BQH method, when combined with PBSA, the QHIC/PBSA actually leads to absolute binding free energies in slightly better agreement with the PMF results (and with experiments) than the BQH/PBSA combination (Table 1). This seemingly puzzling result is because of cancellation of errors in QHIC and implicit solvent PBSA, which is known to introduce error when the molecular nature of solvent is important.[59] Based our recent study,[65] water molecules play important but distinct roles in guest binding with the two host molecules TEMOA and TEETOA, with TEMOA undergoing a dehydration transition whereas guest binding to TEETOA resulted in the opening of the binding cavity that remains essentially dry during the process. It will be challenging for a continuum solvent model such as PBSA or the hybrid 3D-RISM to precisely model the molecular nature of the desolvation processes involved and the errors in the solvation free energy calculated by PBSA could be compensated by the errors in the configurational entropy calculated by QHIC, resulting in a slightly better agreement with the more accurate explicit solvent based PMF results. We believe that the real advantage of BQH over QHIC will be realized when combined with superior solvation free energy methods and applied on more flexible receptor–ligand systems.

Discussion and conclusion

End-point binding free energy methods, while widely used for estimating relative binding strengths (*e.g.* with the widely used MM-PB(GB)/SA), are generally considered less accurate compared with the pathway based methods. As a result, they have rarely been used for estimating absolute binding free energies, except for a few conceptually insightful studies.[21,41,42,60,78] However, despite this generally correct assessment, such methods do have the advantage of being able to decompose the total binding free energy into the physically appealing terms such as configurational entropy that may be related to structural features of the complexed and unbounded state. In this work, we tested the ability of different combinations of solute configurational entropy and solvation free energy models in the end-point approach, including the new EE-BQH method which incorporates the Boltzmann-Quasiharmonic model for configurational entropy in terms of reproducing the absolute binding free energies of the ten host–guest complexes from the SAMPL8 challenge.[65] We found that despite the use of implicit solvent model for solvation, the results obtained with the best combinations QHIC-PBSA and BQH-PBSA are well correlated to those obtained from the more accurate PMF method in explicit solvent (Tables 1 and 2). In terms of quantitatively reproducing the absolute values of the binding free energies the averaged errors in these two methods are about 1.5–1.9 kcal mol^{-1} compared with the theoretical benchmark of the PMF calculation. It is possible that the bulk of the remaining errors in the QHIC-PBSA and BQH-PBSA calculations are to be found in the

implicit solvent model PBSA. Compared with BQH, QHIC underestimates the reduction of configurational entropies in binding because of the ligand rotational fluctuations that are non-Gaussian (Table 4 and Fig. 7). However, because of the cancellation of errors in the QHIC entropies and in the PBSA solvation free energies, the QHIC/PBSA calculated absolute binding free energies are in slightly better agreement with experiments compared with those predicted by BQH/PBSA (Table 1 and Fig. 2). More studies are needed to examine this further by comparing with more accurate explicit solvent based solvation free energy methods.[79–82] The main advantage of the EE-BQH approach outlined is that with this method, the different thermodynamic contributions to the free energy of binding can be decomposed into their physical contributions (translational, rotational, and torsional entropies, correlation entropies, *etc.*). This potentially allows for focusing ligand design efforts by addressing specifically the various contributions to the free energy of binding thereby providing useful guidance for ligand design. Due to the tendency of QHIC to overestimate the first order uncorrelated entropy (Fig. 7 and 8), such information cannot be reliably obtained from QHIC.

Perhaps not unexpectedly, the study shows that the more approximate configurational entropy methods QHCC and NMA lead to significantly poorer correlations with the theoretical benchmark from the PMF calculations and substantially larger errors in the calculated absolute binding free energies (Fig. 3). The results of QHCC and QHIC are consistent with previous studies in which the configurational entropies calculated using QH approximation in the internal coordinates are superior to those using the Cartesian coordinates;[34,36] The main reason is that the degrees of freedom in the Cartesian coordinates are more strongly correlated (*e.g.* three bonded atoms forming a bond angle are highly correlated) than those in the internal coordinates (*e.g.* two adjacent torsions can still move somewhat independently from each other). Such strong correlations in the Cartesian space are difficult to be accounted for using the QH approximation which only considers pairwise linear correlation.

The results of end-point calculations also indicate that for the host–guest systems studied here, the use of 3D-RISM with the linear partial molar volume correction[54,55] for solvation free energy in combination with the different configurational entropy methods are generally less accurate compared with the corresponding combinations using the PBSA continuum solvent model (Tables 1 and 2). While 3D-RISM in combination with QHIC and BQH yield reasonably good correlation with the results from the PMF calculations (with R^2 from 0.48–0.63), the absolute binding free energies are systematically underestimated by 5–6 kcal mol^{-1}, suggesting possible directions for parameter tuning in the 3D-RISM. In summary, the present study highlights the crucial role of both configurational entropy and solvation energy in determining binding free energy and demonstrates the potential of using the new endpoint method to provide insights in more complex protein–ligand systems in combination with more rigorous solvation thermodynamics methods for absolute binding free energy analysis.

Lastly, we note that in recent years, two new configurational entropy methods, Interaction-Entropy method[83] and Cumulant Expansion method[84] that are based on the analysis of fluctuations and higher order expansions of the potential energy respectively, have been developed. These methods are in a sense fundamentally different than the QH type of

approaches studied here that are based on approximating PDF of structural fluctuations. Compared with the structure-based approaches such as BQH, the new energy fluctuation-based approaches are not limited to only linear pairwise correlations along the different degrees of freedom. However, it can be difficult to extract the entropy contributions from different degrees of freedom from the energy-fluctuation based methods. It will be interesting to compare the QH types of structure-based entropy approaches with the energy fluctuation-based approaches in future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dr Kim Sharp for helpful discussions and Dr Julie Harris for reading the manuscript. This study is supported by a Bridge fund from Pace University (N. D.) and National Institutes of Health grant 1R01-GM100946 (T. K.). Part of the calculations are performed on the OWLSNEST high performance cluster at Temple University.

References

1. Heinzelmann G and Gilson MK, Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation, Sci. Rep, 2021, 11(1), 1116. [PubMed: 33441879]
2. Mobley DL and Gilson MK, Predicting binding free energies: Frontiers and benchmarks, Annu. Rev. Biophys, 2017, 46(1), 531–558. [PubMed: 28399632]
3. Yin J, Henriksen NM, Slochower DR and Gilson MK, The SAMPL5 host-guest challenge: Computing binding free energies and enthalpies from explicit solvent simulations by the attach-pull-release (APR) method, J. Comput. Aided Mol. Des, 2017, 31(1), 133–145. [PubMed: 27638809]
4. Chen W, Deng Y, Russell E, Wu Y, Abel R and Wang L, Accurate calculation of relative binding free energies between ligands with different net charges, J. Chem. Theory Comput, 2018, 14(12), 6346–6358. [PubMed: 30375870]
5. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, Murcko M, Frye L, Farid R, Lin T, Mobley DL, Jorgensen WL, Berne BJ, Friesner RA and Abel R, Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field, J. Am. Chem. Soc, 2015, 137(7), 2695–2703. [PubMed: 25625324]
6. Lee TS, Allen BK, Giese TJ, Guo Z, Li P, Lin C, McGee TD Jr., Pearlman DA, Radak BK, Tao Y, Tsai HC, Xu H, Sherman W and York DM, Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery, J. Chem. Inf. Model, 2020, 60(11), 5595–5623. [PubMed: 32936637]
7. He X, Liu S, Lee TS, Ji B, Man VH, York DM and Wang J, Fast, accurate, and reliable protocols for routine calculations of protein-ligand binding affinities in drug design projects using AMBER GPU-TI with ff14SB/GAFF, ACS Omega, 2020, 5(9), 4611–4619. [PubMed: 32175507]
8. Cruz J, Wickstrom L, Yang D, Gallicchio E and Deng N, Combining alchemical transformation with a physical pathway to accelerate absolute binding free energy calculations of charged ligands to enclosed binding sites, J. Chem. Theory Comput, 2020, 16(4), 2803–2813. [PubMed: 32101691]
9. Chaires JB, A thermodynamic signature for drug—DNA binding mode, Arch. Biochem. Biophys, 2006, 453(1), 26–31. [PubMed: 16730635]
10. Mittal S, Bandaranayake RM, King NM, Prabu-Jeyabalan M, Nalam MN, Nalivaika EA, Yilmaz NK and Schiffer CA, Structural and thermodynamic basis of amprenavir/darunavir and atazanavir resistance in HIV-1 protease with mutations at residue 50, J. Virol, 2013, 87(8), 4176–4184. [PubMed: 23365446]

11. Gilson M, Given J, Bush B and McCammon J, The statistical-thermodynamic basis for computation of binding affinities: A critical review, Biophys. J, 1997, 72(3), 1047–1069. [PubMed: 9138555]
12. Boresch S, Tettinger F, Leitgeb M and Karplus M, Absolute binding free energies: A quantitative approach for their calculation, J. Phys. Chem. B, 2003, 107(35), 9535–9551.
13. Wang J, Deng Y and Roux B, Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials, Biophys. J, 2006, 91(8), 2798–2814. [PubMed: 16844742]
14. Woo HJ and Roux B, Calculation of absolute protein—ligand binding free energy from computer simulations, Proc. Natl. Acad. Sci. U. S. A, 2005, 102(19), 6825–6830. [PubMed: 15867154]
15. Nguyen TH and Minh DDL, Implicit ligand theory for relative binding free energies, J. Chem. Phys, 2018, 148(10), 104114. [PubMed: 29544299]
16. Nguyen TH, Zhou HX and Minh DDL, Using the fast fourier transform in binding free energy calculations, J. Comput. Chem, 2018, 39(11), 621–636. [PubMed: 29270990]
17. Minh DD, Implicit ligand theory: Rigorous binding free energies and thermodynamic expectations from molecular docking, J. Chem. Phys, 2012, 137(10), 104106. [PubMed: 22979849]
18. Wu JZ, Azimi S, Khuttan S, Deng N and Gallicchio E, Alchemical transfer approach to absolute binding free energy estimation, J. Chem. Theory Comput, 2021, 17(6), 3309–3319. [PubMed: 33983730]
19. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA and Case DA, Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate—DNA helices, J. Am. Chem. Soc, 1998, 120(37), 9401–9409.
20. Åqvist J, Medina C and Samuelsson J-E, A new method for predicting binding affinity in computer-aided drug design, Protein Eng., Des. Sel, 1994, 7(3), 385–391.
21. Chang CE and Gilson MK, Free energy, entropy, and induced fit in host-guest recognition: Calculations with the second-generation mining minima algorithm, J. Am. Chem. Soc, 2004, 126(40), 13156–13164. [PubMed: 15469315]
22. Deng N-J and Cieplak P, Insights into affinity and specificity in the complexes of α-lytic protease and its inhibitor proteins: Binding free energy from molecular dynamics simulation, Phys. Chem. Chem. Phys, 2009, 11(25), 4968. [PubMed: 19562127]
23. Zhou HX and Gilson MK, Theory of free energy and entropy in noncovalent binding, Chem. Rev, 2009, 109(9), 4092–4107. [PubMed: 19588959]
24. Killian BJ, Yundenfreund Kravitz J and Gilson MK, Extraction of configurational entropy from molecular simulations *via* an expansion approximation, J. Chem. Phys, 2007, 127(2), 024107. [PubMed: 17640119]
25. Suárez D and Díaz N, Direct methods for computing single-molecule entropies from molecular simulations, Wiley Interdiscip. Rev.: Comput. Mol. Sci, 2015, 5(1), 1–26.
26. Chang C, Chen W and Gilson MK, Ligand configurational entropy and protein binding, Proc. Natl. Acad. Sci. U. S. A, 2007, 104(5), 1534–1539. [PubMed: 17242351]
27. Genheden S and Ryde U, The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, Expert Opin. Drug Discovery, 2015, 10(5), 449–461.
28. Hou T, Wang J, Li Y and Wang W, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, J. Chem. Inf Model, 2011, 51(1), 69–82. [PubMed: 21117705]
29. Wang C, Nguyen PH, Pham K, Huynh D, Le TB, Wang H, Ren P and Luo R, Calculating protein—ligand binding affinities with MMPBSA: Method and error analysis, J. Comput. Chem, 2016, 37(27), 2436–2446. [PubMed: 27510546]
30. Wang C, Greene D, Xiao L, Qi R and Luo R, Recent Developments and Applications of the MMPBSA Method, Front. Mol. Biosci, 2017, 4, 87. [PubMed: 29367919]
31. Harpole KW and Sharp KA, Calculation of configurational entropy with a Boltzmann-quasiharmonic model: The origin of high-affinity protein—ligand binding, J. Phys. Chem. B, 2011, 115(30), 9461–9472. [PubMed: 21678965]
32. Miller BR 3rd, McGee TD Jr., Swails JM, Homeyer N, Gohlke H and Roitberg AE, MMPBSA.py: An efficient program for end-state free energy calculations, J. Chem. Theory Comput, 2012, 8(9), 3314–3321. [PubMed: 26605738]

33. Genheden S, Luchko T, Gusarov S, Kovalenko A and Ryde U, An MM/3D-RISM approach for ligand binding affinities, J. Phys. Chem. B, 2010, 114(25), 8505–8516. [PubMed: 20524650]
34. Hikiri S, Yoshidome T and Ikeguchi M, Computational methods for configurational entropy using internal and Cartesian coordinates, J. Chem. Theory Comput, 2016, 12(12), 5990–6000. [PubMed: 27951672]
35. Karplus M and Kushick JN, Method for estimating the configurational entropy of macromolecules, Macromolecules, 1981, 14(2), 325–332.
36. Chang CE, Chen W and Gilson MK, Evaluating the accuracy of the quasiharmonic approximation, J. Chem. Theory Comput, 2005, 1(5), 1017–1028. [PubMed: 26641917]
37. Minh DD, Bui JM, Chang CE, Jain T, Swanson JM and McCammon JA, The entropic cost of protein-protein association: A case study on acetylcholinesterase binding to fasciculin-2, Biophys. J, 2005, 89(4), L25–L27. [PubMed: 16100267]
38. King BM and Tidor B, MIST: Maximum Information Spanning Trees for dimension reduction of biological data sets, Bioinformatics, 2009, 25(9), 1165–1172. [PubMed: 19261718]
39. Cheluvaraja S and Meirovitch H, Calculation of the entropy and free energy of peptides by molecular dynamics simulations using the hypothetical scanning molecular dynamics method, J. Chem. Phys, 2006, 125(2), 24905. [PubMed: 16848609]
40. Hensen U, Grubmuller H and Lange OF, Adaptive anisotropic kernels for nonparametric estimation of absolute configurational entropies in high-dimensional configuration spaces, Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys, 2009, 80(1 Pt 1), 011913.
41. Swanson JMJ, Henchman RH and McCammon JA, Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy, Biophys. J, 2004, 86(1), 67–74. [PubMed: 14695250]
42. Lee MS and Olson MA, Calculation of absolute protein-ligand binding affinity using path and endpoint approaches, Biophys. J, 2006, 90(3), 864–877. [PubMed: 16284269]
43. Deng N-J and Cieplak P, Molecular dynamics and free energy study of the conformational equilibria in the UUUU RNA hairpin, J. Chem. Theory Comput, 2007, 3(4), 1435–1450. [PubMed: 26633215]
44. Di Nola A, Berendsen HJC and Edholm O, Free energy determination of polypeptide conformations generated by molecular dynamics, Macromolecules, 1984, 17(10), 2044–2050.
45. King BM, Silver NW and Tidor B, Efficient calculation of molecular configurational entropies using an information theoretic approximation, J. Phys. Chem. B, 2012, 116(9), 2891–2904. [PubMed: 22229789]
46. Deng N, Cui D, Zhang BW, Xia J, Cruz J and Levy R, Comparing alchemical and physical pathway methods for computing the absolute binding free energy of charged ligands, Phys. Chem. Chem. Phys, 2018, 20(25), 17081–17092. [PubMed: 29896599]
47. Deng N, Wickstrom L, Cieplak P, Lin C and Yang D, Resolving the ligand-binding specificity in c-MYC G-quadruplex DNA: Absolute binding free energy calculations and SPR experiment, J. Phys. Chem. B, 2017, 121(46), 10484–10497. [PubMed: 29086571]
48. Deng N, Xia J, Wickstrom L, Lin C, Wang K, He P, Yin Y and Yang D, Ligand selectivity in the recognition of proto-berberine alkaloids by hybrid-2 human telomeric G-quadruplex: Binding free energy calculation, fluorescence binding, and NMR Experiments, Molecules, 2019, 24(8), 1574.
49. Hirata F, Molecular Theory of Solvation, New York, Boston, Dordrecht, London, Moscow, 2004.
50. Beglov D and Roux B, An integral equation to describe the solvation of polar molecules in liquid water, J. Phys. Chem. B, 1997, 101(39), 7821–7826.
51. Hansen J-P and McDonald IR, Theory of Simple Liquids, Academic Press, 2013.
52. Kovalenko A and Hirata F, Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach, J. Chem. Phys, 2000, 112(23), 10391–10402.
53. Nguyen C, Yamazaki T, Kovalenko A, Case DA, Gilson MK, Kurtzman T and Luchko T, A molecular reconstruction approach to site-based 3D-RISM and comparison to GIST hydration thermodynamic maps in an enzyme active site, PLoS One, 2019, 14(7), e0219473. [PubMed: 31291328]

54. Johnson J, Case DA, Yamazaki T, Gusarov S, Kovalenko A and Luchko T, Small molecule hydration energy and entropy from 3D-RISM, J. Phys.: Condens. Matter, 2016, 28(34), 344002. [PubMed: 27367817]
55. Palmer DS, Frolov AI, Ratkova EL and Fedorov MV, Towards a universal method for calculating hydration free energies: A 3D reference interaction site model with partial molar volume correction, J. Phys. Condens. Matter, 2010, 22(49), 492101. [PubMed: 21406779]
56. Wickstrom L, Deng N, He P, Mentes A, Nguyen C, Gilson MK, Kurtzman T, Gallicchio E and Levy RM, Parameterization of an effective potential for protein—ligand binding from host—guest affinity data: Force field optimization with host—guest systems, J. Mol. Recognit, 2016, 29(1), 10–21. [PubMed: 26256816]
57. Wickstrom L, He P, Gallicchio E and Levy RM, Large scale affinity calculations of cyclodextrin host—guest complexes: Understanding the role of reorganization in the molecular recognition process, J. Chem. Theory Comput, 2013, 9(7), 3136–3150. [PubMed: 25147485]
58. Harris RC, Deng N, Levy RM, Ishizuka R and Matubayasi N, Computing conformational free energy differences in explicit solvent: An efficient thermodynamic cycle using an auxiliary potential and a free energy functional constructed from the end points, J. Comput. Chem, 2017, 38(15), 1198–1208. [PubMed: 28008630]
59. He P, Sarkar S, Gallicchio E, Kurtzman T and Wickstrom L, Role of displacing confined solvent in the conformational equilibrium of beta-cyclodextrin, J. Phys. Chem. B, 2019, 123(40), 8378–8386. [PubMed: 31509409]
60. Tang Z and Chang CA, Binding thermodynamics and kinetics calculations using chemical host and guest: A comprehensive picture of molecular recognition, J. Chem. Theory Comput, 2018, 14(1), 303–318. [PubMed: 29149564]
61. Ali HS, Chakravorty A, Kalayan J, de Visser SP and Henchman RH, Energy—entropy method using multiscale cell correlation to calculate binding free energies in the SAMPL8 host-guest challenge, J. Comput. Aided Mol. Des, 2021, 35(8), 911–921. [PubMed: 34264476]
62. SAMPL8 Challenge GDCC set, https://github.com/samplchallenges/SAMPL8/tree/master/host_guest/GDCC.
63. Gumbart JC, Roux B and Chipot C, Standard binding free energies from computer simulations: What is the best strategy?, J. Chem. Theory Comput, 2013, 9(1), 794–802. [PubMed: 23794960]
64. Gumbart JC, Roux B and Chipot C, Efficient determination of protein—protein standard binding free energies from first principles, J. Chem. Theory Comput, 2013, 9(8), 3789–3798.
65. Azimi S, Wu JZ, Khuttan S, Kurtzman T, Deng N and Gallicchio E, Application of the alchemical transfer and potential of mean force methods to the SAMPL8 host-guest blinded challenge, J. Comput.-Aided Mol. Des, 2022, 36, 63–76. [PubMed: 35059940]
66. Gallicchio E, Free Energy-Based Computational Methods for the Study of Protein-Peptide Binding Equilibria, Springer Nature, 2021.
67. Hill TL, Cooperativity Theory in Biochemistry, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1985.
68. Widom B, Some topics in the theory of fluids, J. Chem. Phys, 1963, 39(11), 2808–2812.
69. Fenley AT, Henriksen NM, Muddana HS and Gilson MK, Bridging calorimetry and simulation through precise calculations of cucurbituril—guest binding enthalpies, J. Chem. Theory Comput, 2014, 10(9), 4069–4078. [PubMed: 25221445]
70. Luchko T, Gusarov S, Roe DR, Simmerling C, Case DA, Tuszynski J and Kovalenko A, Three-dimensional molecular theory of solvation coupled with molecular dynamics in Amber, J. Chem. Theory Comput, 2010, 6(3), 607–624. [PubMed: 20440377]
71. Chang CE, McLaughlin WA, Baron R, Wang W and McCammon JA, Entropic contributions and the influence of the hydrophobic environment in promiscuous protein—protein association, Proc. Natl. Acad. Sci. U. S. A , 2008, 105(21), 7456–7461. [PubMed: 18495919]
72. Tidor B and Karplus M, The contribution of vibrational entropy to molecular association. The dimerization of insulin, J. Mol. Biol, 1994, 238(3), 405–414. [PubMed: 8176732]
73. McQuarrie DA, Statistical Mechanics, University Science Books, 2000.
74. Case DA, Belfon K, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE III, Cruzeiro VWD, Darden TA, Duke RE, Giambasu G, Gilson MK, Gohlke H, Goetz AW, Harris R, Izadi S, Izmailov

SA, Kasavajhala K, Kovalenko A, Krasny R, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Man V, Merz KM, Miao Y, Mikhailovskii O, Monard G, Nguyen H, Onufriev A, Pan F, Pantano S, Qi R, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Skrynnikov NR, Smith J, Swails J, Walker RC, Wang J, Wilson L, Wolf RM, Wu X, Xiong Y, Xue Y, York DM and Kollman PA, AMBER 2020, University of California, San Francisco, 2020.

75. Wang J, Wolf RM, Caldwell JW, Kollman PA and Case DA, Development and testing of a general amber force field, J. Comput. Chem, 2004, 25(9), 1157–1174. [PubMed: 15116359]
76. Jakalian A, Bush BL, Jack DB and Bayly CI, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method, J. Comput. Chem, 2000, 21(2), 132–146.
77. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H and Pedersen LG, A smooth particle mesh Ewald method, J. Chem. Phys, 1995, 103(19), 8577.
78. Huang Y.-m. M., Chen W, Potter MJ and Chang C-A, Insights from free-energy calculations: Protein conformational equilibrium, driving forces, and ligand-binding modes, Biophys. J, 2012, 103(2), 342–351. [PubMed: 22853912]
79. Chen L, Cruz A, Roe DR, Simmonett AC, Wickstrom L, Deng N and Kurtzman T, Thermodynamic decomposition of solvation free energies with particle mesh ewald and long-range Lennard-Jones interactions in grid inhomogeneous solvation theory, J. Chem. Theory Comput, 2021, 17(5), 2714–2724. [PubMed: 33830762]
80. Ramsey S, Nguyen C, Salomon-Ferrer R, Walker RC, Gilson MK and Kurtzman T, Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST, J. Comput. Chem, 2016, 37(21), 2029–2037. [PubMed: 27317094]
81. Nguyen CN, Young TK and Gilson MK, Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril, J. Chem. Phys, 2012, 137(4), 044101. [PubMed: 22852591]
82. Waibl F, Kraml J, Fernandez-Quintero ML, Loeffler JR and Liedl KR, Explicit solvation thermodynamics in ionic solution: Extending grid inhomogeneous solvation theory to solvation free energy of salt-water mixtures, J. Comput. Aided Mol. Des, 2022, DOI: 10.1007/s10822-021-00429-y.
83. Duan L, Liu X and Zhang JZ, Interaction entropy: A new paradigm for highly efficient and reliable computation of protein-ligand binding free energy, J. Am. Chem. Soc, 2016, 138(17), 5722–5728. [PubMed: 27058988]
84. Menzer WM, Li C, Sun W, Xie B and Minh DDL, Simple entropy terms for end-point binding free energy calculations, J. Chem. Theory Comput, 2018, 14(11), 6035–6049. [PubMed: 30296084]

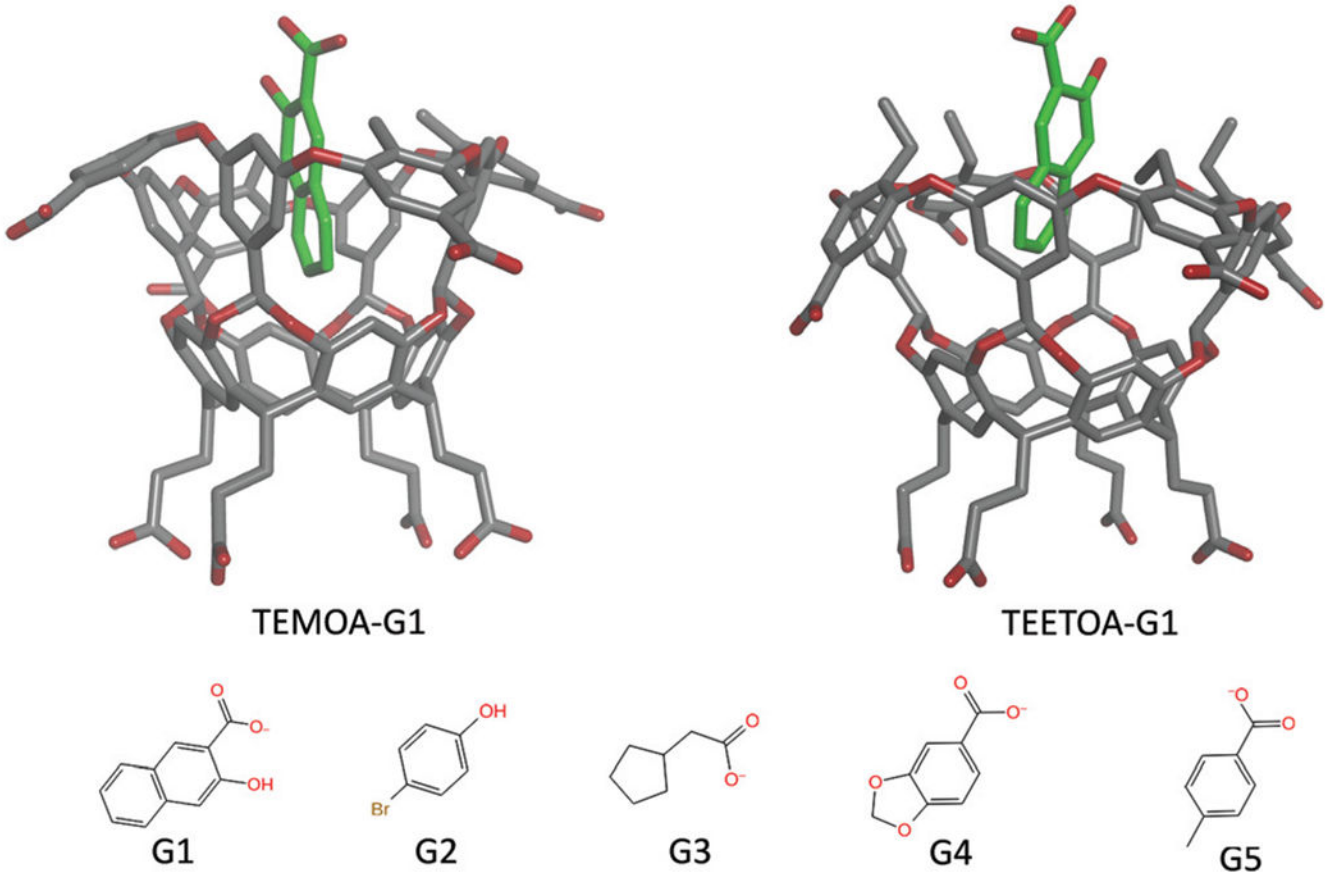

Fig. 1.

Representative MD structures of TEMOA-G1 and TEETOA-G1 complexes and chemical structures of the guest molecules.

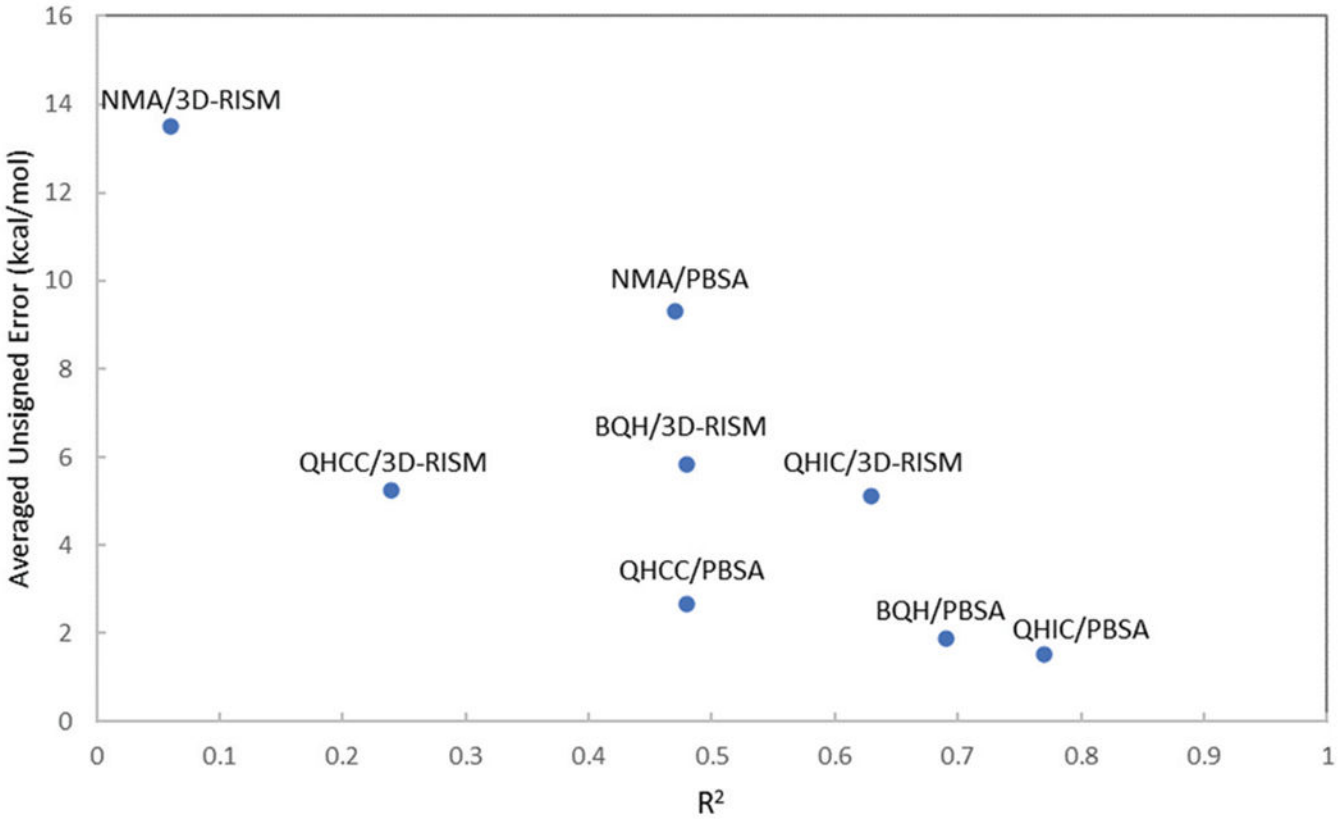

Fig. 2.

Average deviations and R^2 between the end-point binding free energy estimates using different combinations of configurational entropy and solvation free energy treatments and the reference PMF results.

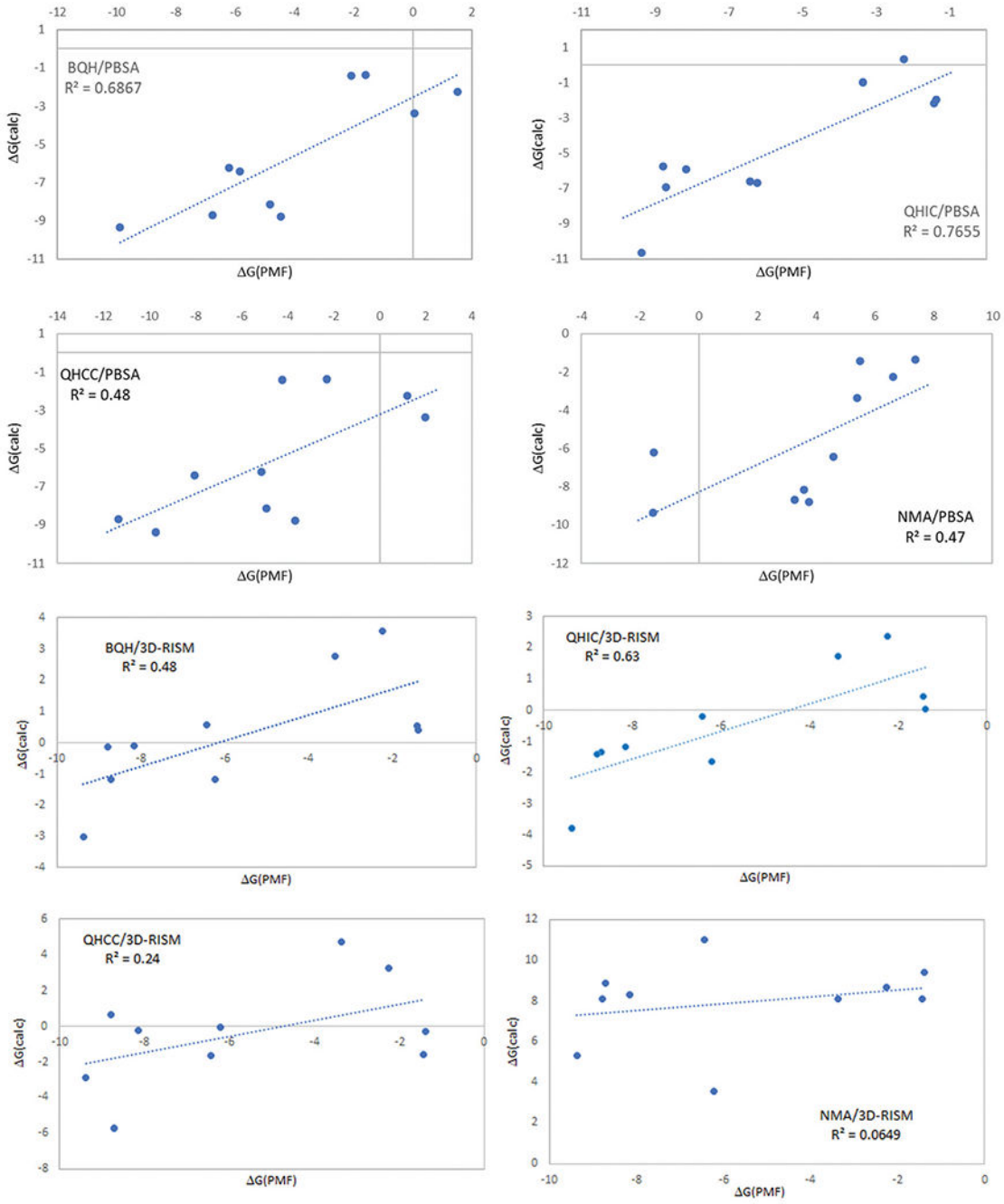

Fig. 3.

Correlations of the absolute binding free energies from the end-point calculations with the PMF results.

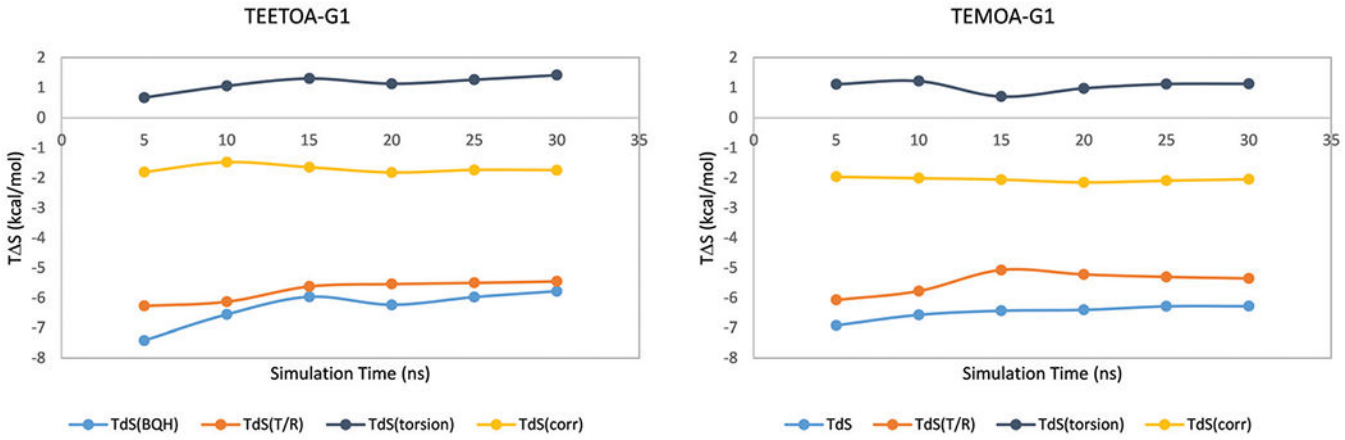

Fig. 4.

The cumulative BQH entropy and its components during the simulations for the TEETOA-G1 and TEMOA-G1 systems.

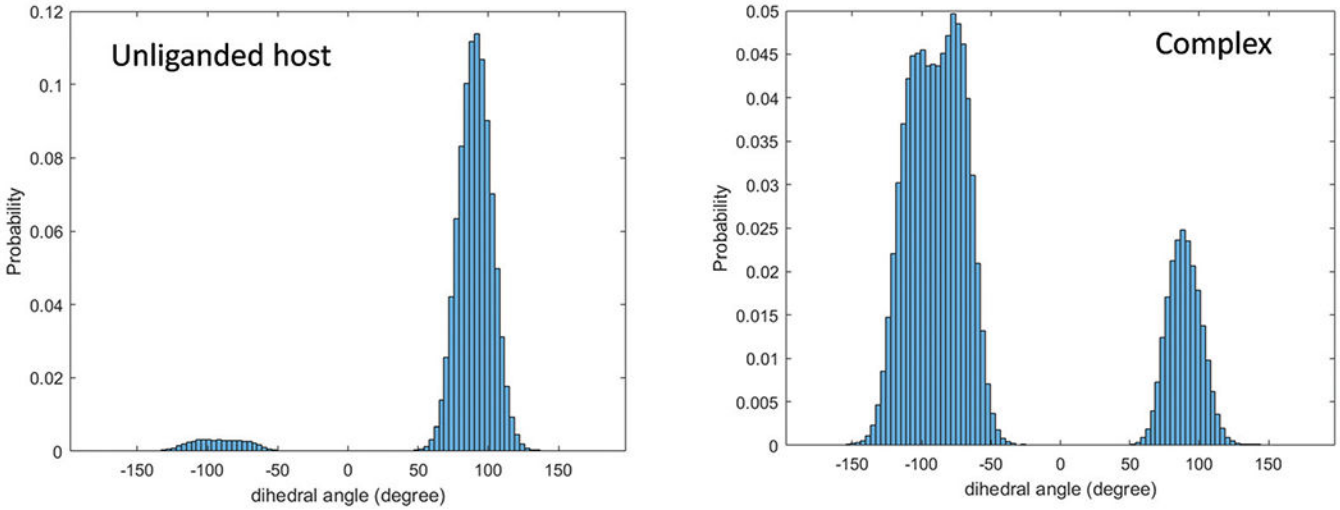

Fig. 5.

The histograms of the torsion angle distribution for one of the ethyl sidechains in the unliganded TEETOA host and that in the TEETOA-G1 complex sampled during a 30 ns MD in solution.

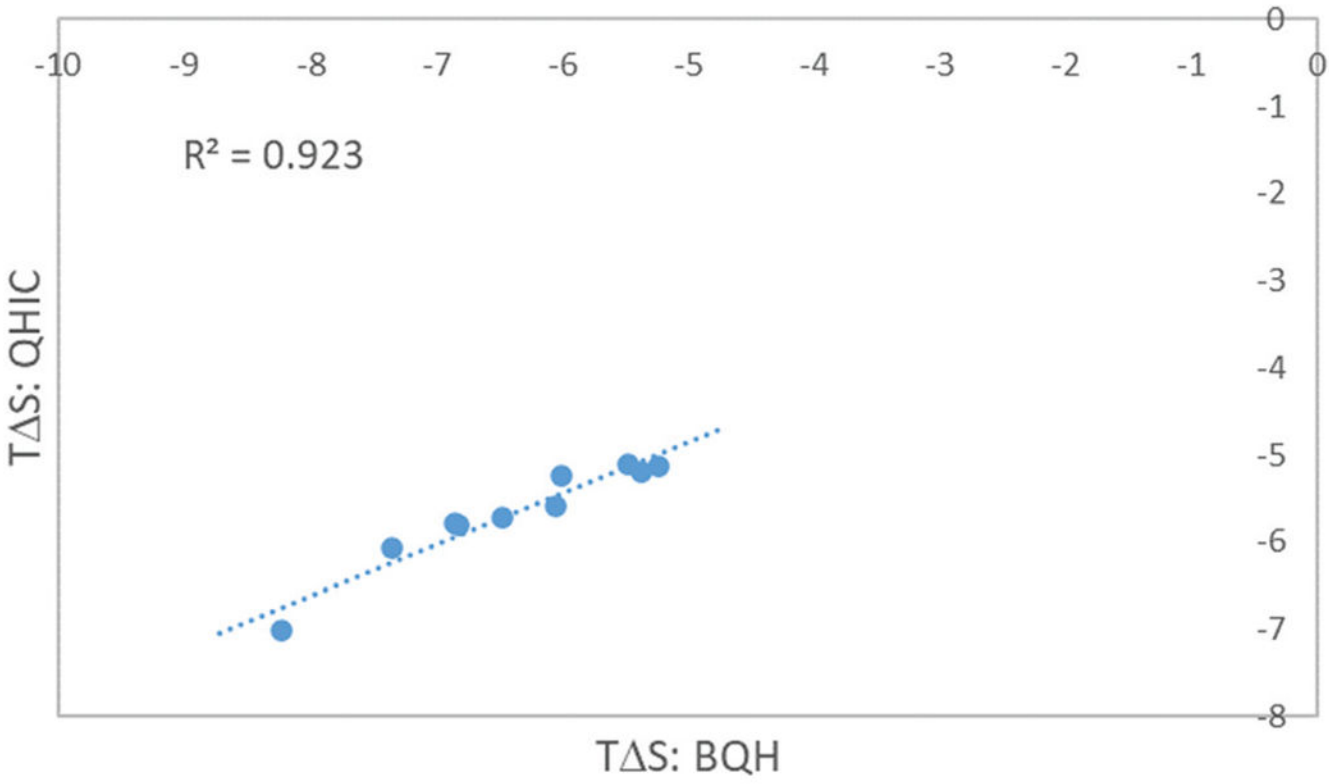

Fig. 6.

The correlation between the configurational entropy of binding $T\Delta S^{\circ}_{\text{config}}$ computed using BQH and QHIC.

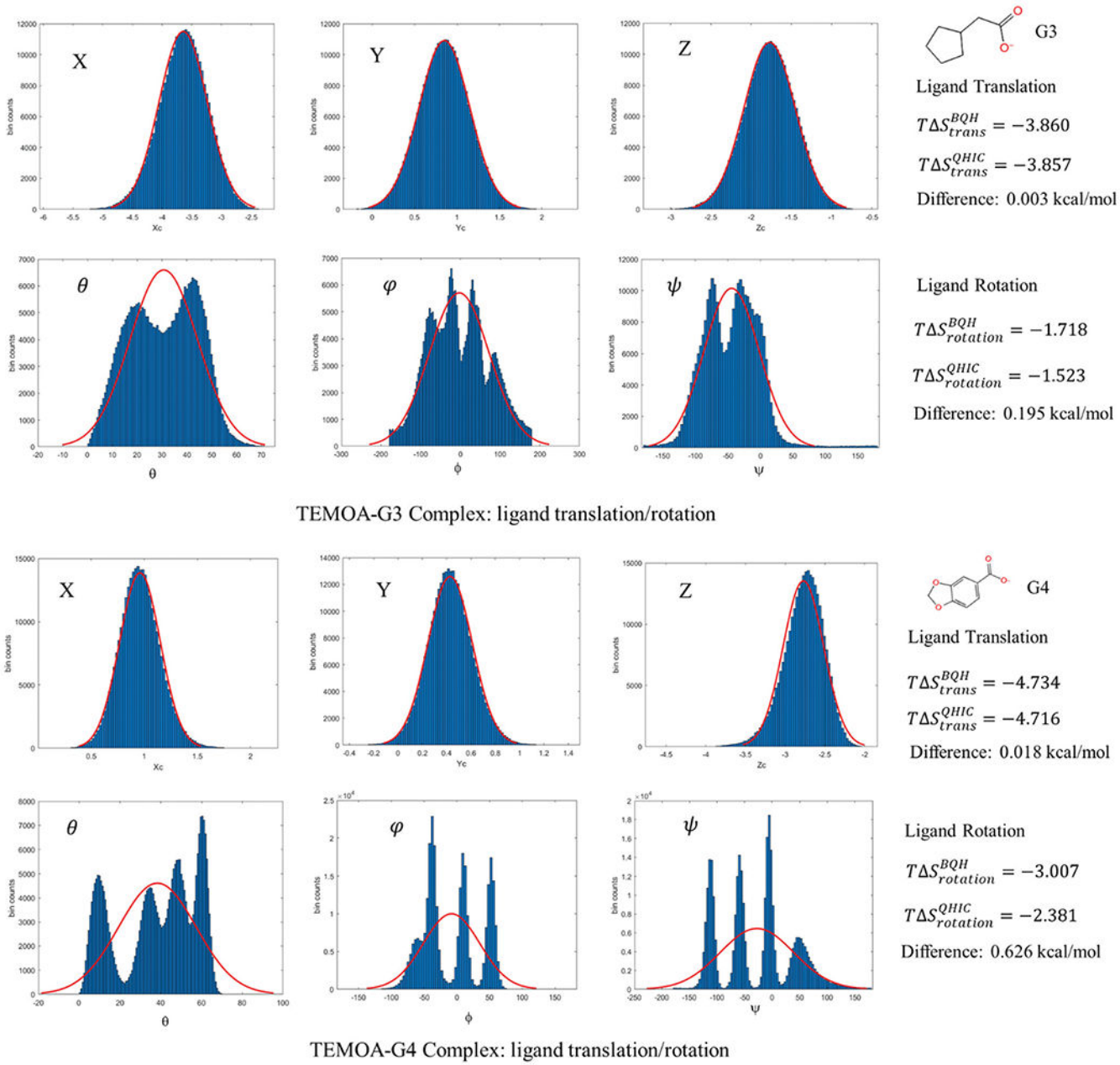

Fig. 7.

Histograms of the distributions along the three translational and three rotation degrees of freedom (Euler angles θ, φ, and ψ) of the G3 guest molecules in the TEMOA-G3 and TEMOA-G4 complexes. The red curves are the single Gaussian fits. The translational and rotational entropy losses upon binding are calculated as $\Delta S_{\text{trans}} = S_{\text{trans}} - k \ln V^{\circ}$ and $\Delta S_{\text{rotation}} = S_{\text{rotation}} - k \ln 8\pi^2$.

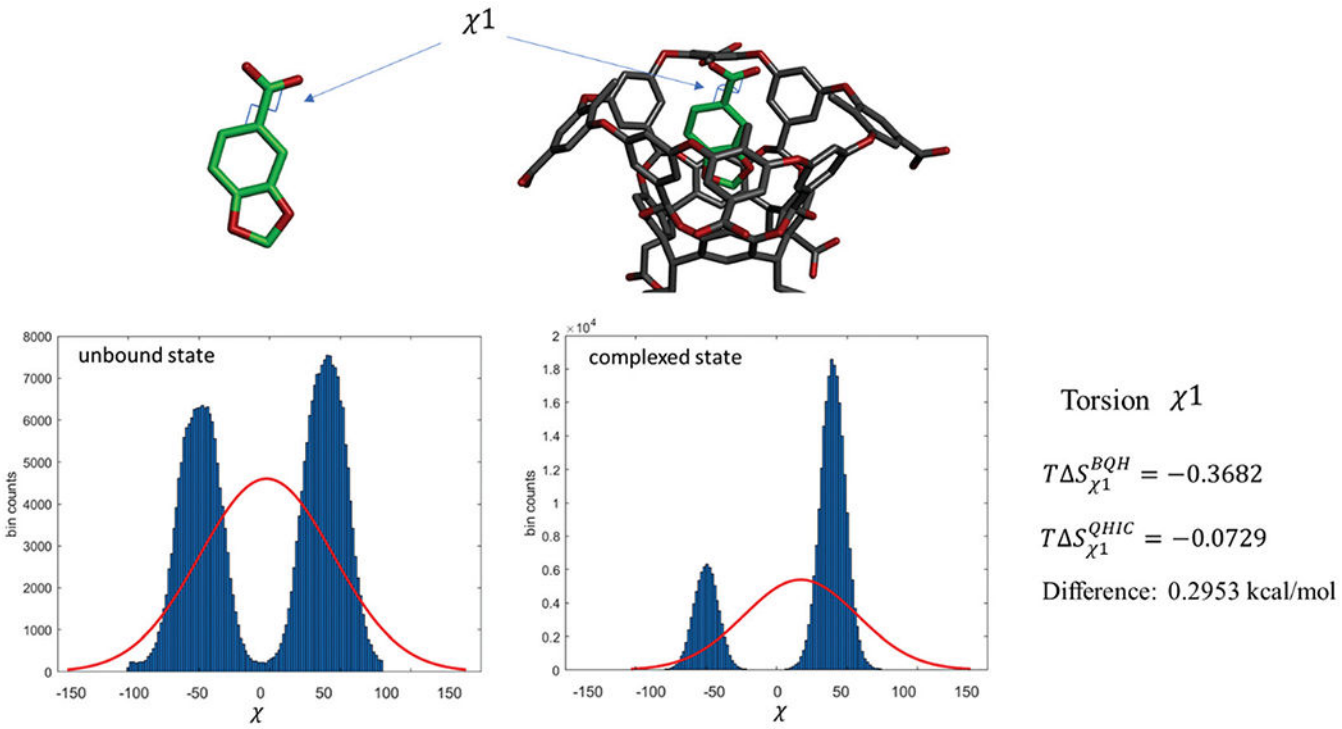

Fig. 8.

The distributions of the torsion angle $\chi 1$ in the guest molecule G3 in the TEMOA-G3 complex and unbound state. The red curves are the single Gaussian fits. The torsional entropy loss $\Delta S_{\chi 1} = S_{\chi 1}$ (complexed) − $S_{\chi 1}$ (unbound) calculated by BQH and QHIC are also shown.

Table 1

R^2 and averaged unsigned errors (AUE) of the absolute binding free energy estimates from the end-point calculations relative to the results of the PMF method[62]

	BQH/ PBSA	QHIC/ PBSA	QHCC/ PBSA	NMA/ PBSA	BQH/3D-RISM	QHIC/3D-RISM	QHCC/3D-RISM	NMA/3D-RISM
R^2	0.69	0.77	0.48	0.47	0.48	0.63	0.24	0.06
AUE[a]	1.88	1.53	2.65	9.30	5.84	5.11	5.26	13.5

[a] In kcal mol^{-1}.

Table 2

Absolute binding free energies in kcal mol^{-1} for the SAMPL8 GDCC host–guest complexes calculated using the different end-point methods and the PMF method

Host–guest	BQH/PBSA	QHIC/PBSA	QHCC/PBSA	NMA/PBSA	BQH/3D-RISM	QHIC/3D-RISM	QHCC/3D-RISM	NMA/3D-RISM	PMF
TEETOA-G1	−1.60	−1.96	−2.31	7.37	0.42	0.05	−0.30	9.38	−1.38
TEETOA-G2	−6.21	−6.68	−5.14	−1.53	−1.15	−1.62	−0.09	3.53	−6.22
TEETOA-G3	−2.07	−2.18	−4.23	5.49	0.55	0.44	−1.61	8.11	−1.42
TEETOA-G4	1.52	0.32	1.18	6.61	3.57	2.37	3.23	8.66	−2.25
TEETOA-G5	0.05	−0.98	1.98	5.40	2.77	1.74	4.70	8.12	−3.36
TEMOA-G1	−5.85	−6.62	−8.05	4.57	0.57	−0.20	−1.63	10.99	−6.43
TEMOA-G2	−9.89	−10.64	−9.72	−1.57	−3.03	−3.79	−2.86	5.29	−9.37
TEMOA-G3	−6.76	−6.94	−11.34	3.26	−1.15	−1.33	−5.73	8.87	−8.71
TEMOA-G4	−4.47	−5.75	−3.67	3.76	−0.13	−1.41	0.67	8.10	−8.79
TEMOA-G5	−4.83	−5.90	−4.94	3.59	−0.10	−1.18	−0.21	8.32	−8.15

Table 3

The configurational entropy contributions to binding calculated from BQH[ab]

System	$T\Delta S_1^{trans}$	$T\Delta S_1^{rot}$	$T\Delta S_1^{torsion}$	$T\Delta S_1$	$T\Delta S_2^{corr}$	$T\Delta S_{config}^{\circ}$
TEETOA-G1	−3.43	−2.02	1.41	−4.04	−1.43	−5.47
TEETOA-G2	−4.35	−2.05	1.46	−4.94	−1.11	−6.05
TEETOA-G3	−3.43	−2.00	1.67	−3.77	−1.47	−5.24
TEETOA-G4	−4.15	−2.56	0.34	−6.37	−1.85	−8.22
TEETOA-G5	−3.86	−2.11	0.81	−5.16	−1.66	−6.83
TEMOA-G1	−3.51	−1.85	1.13	−4.23	−1.78	−6.01
TEMOA-G2	−4.30	−2.00	1.32	−4.97	−1.50	−6.48
TEMOA-G3	−3.82	−1.55	0.31	−5.07	−0.31	−5.37
TEMOA-G4	−4.71	−2.01	1.18	−5.54	−1.81	−7.35
TEMOA-G5	−4.14	−2.12	1.06	−5.20	−1.65	−6.85

[a] T = 300 K. Unit: kcal mol^{-1}.

[b] $\Delta S_{config}^{\circ} = \Delta S_1 + \Delta S_2^{corr}$, $\Delta S_1 = \Delta S_1^{trans} + \Delta S_1^{rot} + \Delta S_1^{torsion}$.

Table 4

Comparisons of the configurational entropy change in binding $T\Delta S^{\circ}_{\text{config}}$ calculated using BQH and QHIC (unit: kcal mol^{-1})

System	$T\Delta S^{\text{BQH}}_{\text{config}}$	$T\Delta S^{\text{QHIC}}_{\text{config}}$	$T(\Delta S^{\text{QHIC}}_{\text{config}} - S^{\text{BQH}}_{\text{config}})$
TEETOA-G1	−5.47	−5.11	0.36
TEETOA-G2	−6.05	−5.58	0.47
TEETOA-G3	−5.24	−5.13	0.11
TEETOA-G4	−8.22	−7.02	1.21
TEETOA-G5	−6.83	−5.80	1.03
TEMOA-G1	−6.01	−5.24	0.78
TEMOA-G2	−6.48	−5.73	0.75
TEMOA-G3	−5.37	−5.19	0.18
TEMOA-G4	−7.35	−6.07	1.28
TEMOA-G5	−6.85	−5.78	1.07