This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Evaluation of Generalized Born Models for Large Scale Affinity Prediction of Cyclodextrin Host−Guest Complexes

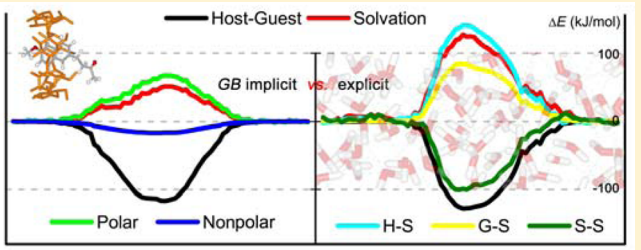
Haiyang Zhang,[†] Chunhua Yin,[†] Hai Yan,[*,†] and David van der Spoel[*,‡]

[†]Department of Biological Science and Engineering, School of Chemistry and Biological Engineering, University of Science and Technology Beijing, 100083 Beijing, China

[‡]Uppsala Center for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, SE-75124 Uppsala, Sweden

Supporting Information

ABSTRACT: Binding affinity prediction with implicit solvent models remains a challenge in virtual screening for drug discovery. In order to assess the predictive power of implicit solvent models in docking techniques with Amber scoring, three generalized Born models (GB^{HCT}, $GB^{OBC}I$, and $GB^{OBC}II$) available in Dock 6.7 were utilized, for determining the binding affinity of a large set of β-cyclodextrin complexes with 75 neutral guest molecules. The results were compared to potential of mean force (PMF) free energy calculations with four GB models (GB^{Still}, GB^{HCT}, $GB^{OBC}I$, and $GB^{OBC}II$) and to experimental data. Docking results yield similar accuracy to the computationally demanding PMF method with umbrella sampling. Neither docking nor PMF calculations reproduce the experimental binding affinities, however, as indicated by a small Spearman rank order coefficient (~0.5). The binding energies obtained from GB models were decomposed further into individual contributions of the binding partners and solvent environments and compared to explicit solvent simulations for five complexes allowing for rationalizing the difference between explicit and implicit solvent models. An important observation is that the explicit solvent screens the interaction between host and guest much stronger than GB models. In contrast, the screening in GB models is too strong in solutes, leading to overestimation of short-range interactions and too strong binding. It is difficult to envision a way of overcoming these two opposite effects.



1. INTRODUCTION

Fast and accurate evaluation of binding affinity for small molecules interacting with specific biological targets is highly imperative in modern medicinal chemistry and biochemistry, but it remains challenging to obtain accurate results. Virtual screening acts as a crucial complement to high-throughput screening by providing, in theory, a cost-effective strategy to generate starting points for drug design and for understanding the mechanism underlying particular biological processes.[1] Due to the high computational cost for evaluating the solvation contributions from an atomic-level description of solvent molecules, simplified treatments such as implicit solvent models are often used instead in virtual screening to evaluate binding affinities of protein−ligand complexes.[2] Most implicit models treat the solvent as a continuous medium with certain dielectric and interfacial properties. Such treatment has inevitable deficiencies in, for instance, short-range interactions and solvent orientation (e.g., solute−solvent hydrogen bonds),[3−5] and further improvements of implicit solvent models are required.[6]

The Generalized Born (GB) model is one of the most popular implicit solvent models for bimolecular simulations and molecular design.[6] Compared to the, in principle, more accurate Poisson−Boltzmann (PB) theory,[7] GB-based calculations were shown to have a similar performance for prediction of solute−solvent electrostatic polarizations in some cases,[8−11] allowing for modeling electrostatics (ΔE_{pol}) in a reasonably accurate and cost-effective manner. Note that whether PB is more accurate than GB or not is a matter of debate, as indicated, for instance, by the comparisons between GB and PB for solvation free energies and the SAMPL3 challenge.[12,13] The nonpolar part (ΔE_{np}) in GB models is typically derived from total solvent-accessible surface areas (SA) scaled by surface tension. The combination of a molecular mechanics (MM) force field and a GB solvation term (giving the MM/GBSA approach) has therefore become popular for scoring protein−ligand associations in large-scale virtual screening.[14−16] For instance, Amber molecular mechanics with GB/SA solvation (denoted as Amber GB/SA) scoring were reported to improve accuracy of scoring docked poses for RNA-ligand complexes,[17] and this method has aided in identifying a lead compound for blocking the growth of cancer cells.[18]

A standard MM/GBSA analysis is performed on MD snapshots of binding partners extracted from explicit simulations, and then binding free energies are evaluated from

Received: July 18, 2016
Published: September 14, 2016

© 2016 American Chemical Society

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

$$\Delta G_{bind} = \Delta E_{MM} + \Delta E_{pol} + \Delta E_{np} - T\Delta S \quad (1)$$

where ΔE_{MM} is the total gas phase energy including bonded and nonbonded interactions, ΔE_{pol} and ΔE_{np} are polar and nonpolar contributions from GB solvent models mentioned above, and ΔS is the configurational entropy.[19,20] Such an analysis typically does not include receptor flexibility, since the same snapshots are used for both the complex and the free receptor. Amber GB/SA scoring (a variant of MM/GBSA) follows the same scheme (eq 1) to evaluate binding affinities (ΔS is not included yet in the Amber score of Dock 6.7) but uses separate GB implicit simulation trajectories for receptor, ligand, and the corresponding complex in order to account for flexibility of the binding partners.[21]

A range of publications has highlighted the complexity and drawbacks of binding affinity predictions based on the MM/GBSA scoring, and the predictions were shown to depend on various parameters such as charge and solvent model, internal dielectric constants, and conformational reorganization.[22−25] Single-conformations are usually not a good representation of conformational ensembles, in particular for flexible and dynamic binding in biomolecules, and therefore implicit solvation free energies, based on a single structure, are likely unreliable.[5,26] By enabling receptor flexibility in whole or in part via implicit solvation molecular dynamics (MD) simulations, GB/SA scorings could, in principle, improve, by modeling the so-called "induced-fit" effects upon association.

Here we focus on the configurational flexibility of binding partners by umbrella sampling and aim to evaluate the performance of Amber GB/SA scoring for the prediction of binding affinities. Beta-cyclodextrin (β-CD) was chosen as a model host, and its binding affinities with a set of 75 guest molecules are evaluated. Cyclodextrins are ideal receptor models for investigating host−guest recognition processes[25,27,28] and for benchmarking the accuracy of GB solvation energy predictions,[23,29,30] since a large amount of experimental reference data is available.[31−35] Potential of mean forces (PMFs) for the formation process of [β-CD:guest] complexes were computed with umbrella sampling,[36] allowing determination of binding (free) energies. Such bound complexes were evaluated by Amber GB/SA scoring in DOCK6.7[21] as well, and the binding affinity and binding modes were then compared with the PMF calculations. Following our previous work on the energy decomposition in explicit solvents,[37] a decomposition for GB models is proposed, highlighting the difference in solvation treatment between the solvent models in detail. Lack of solvent balance and unphysical Coulomb screening in GB models are discussed in the end, and we believe this work may be valuable for evaluating biotechnology applications such as pharmaceutical screening.

2. MATERIALS AND METHODS

2.1. Guest Selection. Thermodynamic data (ΔG^0, ΔH^0, and ΔS^0) for β-CD binding with 75 neutral guest molecules were extracted from a compilation of the experimental determinations by the groups of Inoue and Rekharsky.[31−35] These selected guests can be parametrized properly by the generalized Amber force field;[38] that is, the "antechamber" tool[39] did not give any warnings during parametrization. Only the *R*-stereoisomer conformation of guests was considered here (note that the calculations in this work use the same isomers as the experiments), and we also included some guests studied by the Gilson[40] and Levy[25] groups in case the ΔH^0 and ΔS^0 data are available. The guests selected have diverse topologies including (cyclo)alkane, aromatic, heterocyclic, naphthalene, and sulfonyl groups, covering most of the common elements (C, H, O, N, S, F, Cl, Br) in (halo)organic molecules. Chemical structures of 17 representative guests are shown in Figure 1. A complete list of the 75 studied guests, about 30% of which are featured in the DrugBank database,[41] is given in Scheme S1 and Table S1 in the Supporting Information.

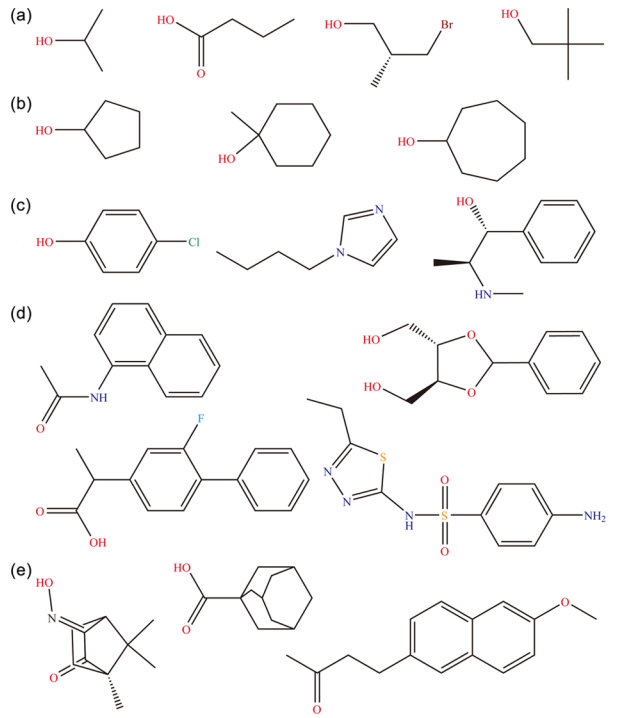


Figure 1. Chemical structures of 17 guest molecules with representative groups in the set of 75 tested guests. (a) Nonring molecules with alkane chains (this class contains 12 members), (b) monocyclic cycloalkanes (8 members), (c) with one aromatic or monocyclic heterocyclic ring (25 members), (d) with two or more rings (27 members), and (e) common strong binders of (*R*)-camphorquinone-3-oxime, 1-adamantanecarboxylic acid, and nabumetone. All the 2-D structures were rotated such that a more hydrophobic moiety of guest lies on the right. A complete list of tested guests is given in Scheme S1 in the Supporting Information.

2.2. Interaction Force Field. 3D conformers of all the tested guests were extracted from the PubChem databases in the Structure Data Format (SDF).[42] With the aid of "antechamber" package,[39] restrained electrostatic potential (RESP) charges of guest molecules were computed via fitting partial charges to electrostatic potentials calculated using the Gaussian 03 software[43] at the Hartree−Fock level with the 6-31G* basis set[44] and using the Merz−Singh−Kollman (MK) scheme.[45] Topology parameters of guest molecules were then generated automatically using the generalized Amber force field (GAFF)[38] by antechamber. Details of the GAFF parametrization for organic molecules have been presented in refs 46 and 47. The initial coordinates of the host β-CD were taken from the crystal structure (PDB code: 1DMB),[48] and the q4md-CD force field[49] was used to model β-CD. As an Amber-like force field, q4md-CD is consistent with GAFF[38] and has shown good performance for modeling cyclodextrin-based systems.[28,49]

All three GB models supported in the GROMACS package,[50] namely, GB^{Still},[51] GB^{HCT},[52] and GB^{OBC} with two

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

variants ($GB^{OBC}I$ and $GB^{OBC}II$),[8] were applied to model the continuous solvent media with a dielectric constant of 80 (corresponding to water as a solvent). The set of van der Waals radii used for evaluation of effective Born radii was taken from the TINKER[53] implementation of GB implicit solvation models. All four GB modes share the same radii in TINKER,[53] while the "mbondi" radii are recommended for use with GB^{HCT} and the "mbondi2" radii are for the two GB^{OBC} variants in the AMBER toolkit.[54] The so-called overlap scaling factors for GB^{HCT} and GB^{OBC} in AMBER were taken from TINKER directly. These GB radii and scaling factors are given in Table S2 in the Supporting Information. Following Mobley's work,[55] we tested the solvation free energies with different radii against the FreeSolv database (latest version 0.32 with 643 solutes).[56] The four GB modes with the radii from TINKER give a correlation of R = ~0.84 between calculated and experimental hydration free energies, in excellent agreement with Mobley's report[55] (R = 0.84 for an old version with 504 solutes) and also with the tests of GB^{HCT} using the "mbondi" radii (R = 0.84) and GB^{OBC} with the "mbondi2" radii (R = 0.82), as shown in Table S3 in the Supporting Information. For consistency with GB^{Still}, the GB radii from TINKER were used in the following calculations.

2.3. Umbrella Sampling. The host β-CD was placed in the center of a cubic box ($5 \times 5 \times 5$ nm^3) with the CD cavity axis parallel to the Z-axis and the secondary rim of β-CD pointing to the $-Z$ direction, as shown in Figure 2. The guest was oriented

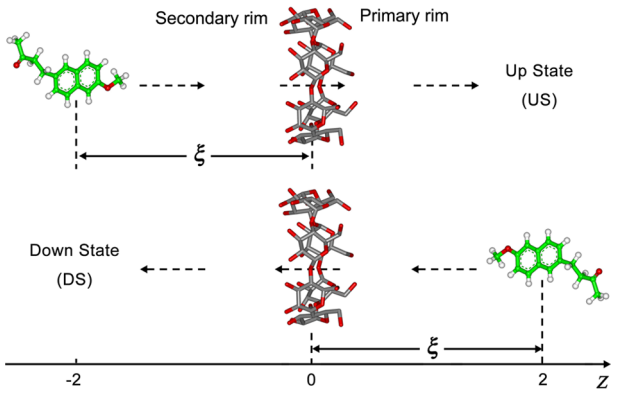


Figure 2. Definition of the reaction coordinate (ξ) and binding modes (US/DS) for β-CD host–guest complexes.

to have its longest axis parallel to the Z-axis and then added to the center of the simulated box, forming an inclusion complex with β-CD. Subsequently, the guest was moved for 40 steps of 0.05 nm, along the $-Z$ and $+Z$ directions, respectively. The resulting 81 configurations, covering a range of [−2, 2] nm, mimick the formation process of the [β-CD:guest] complex.

Due to the structural asymmetry of CDs, β-CD host–guest complexes can be arranged in two binding modes. Here we assumed that the more hydrophobic moiety of guest (e.g., the naphthalene ring of nabumetone in Figure 2) penetrating into the CD cavity from the secondary rim of β-CD along the $+Z$ direction forms an up-state (US) complex, while the penetration from the primary rim of β-CD along $-Z$ gives a down-state (DS) complex. For illustration of the binding modes, the more hydrophobic moiety of guest molecules was placed on the right-hand side in Figure 1 and Scheme S1. In general, aromatic rings are assumed to be the most hydrophobic, followed by heterocyclic rings and then cycloalkanes. Hydrophobicity scales of small functional groups are on the order of $-CH_3$ > $-Br$ > $-Cl$ > $-O-C(=O)CH_3$ > $-C(=O)OH$ > $-C(=O)CH_3$ > $-NH_2$ > $-OH$, based on the experimental Log$P$ values[42] (Table S4) of X$-CH_3$ compounds where X stands for the functional groups above.

Following the same approach, binding processes in the US and DS states for all the 75 guests were generated and were used for subsequent umbrella sampling simulations with the GROMACS software (version 4.5.5).[57−59] All the simulations were carried out without periodic boundary conditions and without cutoffs (all-vs-all) for electrostatic and van der Waals interactions. The temperature was coupled to the value at which β-CD binding free energy data were measured experimentally using the velocity rescaling thermostat with a coupling constant of 0.1 ps.[60] Pressure coupling was turned off because of the vacuum boundary conditions. All bond lengths of β-CD and guest molecules were constrained by the LINCS algorithm,[61] allowing an integration time step of 0.002 ps.

The generated configurations were used as 81 successive windows for construction of free energy profiles (i.e, PMFs) of β-CD binding reactions with guest molecules. The center of mass (COM) distance along Z-axis was defined to be the reaction coordinate ξ (Figure 2), and the distance between host and guest in each window was restrained during the simulations with a harmonic force constant of 2000 kJ mol^{-1} nm^{-2}. After energy minimization, each window was simulated for 20 ns, which was enough for the convergence of the PMF profiles (Figure S1). Since we have four GB models and 75 guests binding to β-CD in two modes (US/DS), a total of 600 PMF profiles were constructed with a total simulation time of 972 μs. Following our previous work,[37] explicit PMF simulations for the [β-CD:2-propanol] complex in the DS mode (81 umbrella windows and 10 ns simulation for each window) were performed as well. Together with the explicit simulations of β-CD complexes with nabumetone (already included in the tested guests) and three isoflavone analogues,[37] these five cases were used to compare to GB simulations for exploring the difference in explicit and implicit GB solvent models. GB PMF simulations of the three isoflavones (daidzein, daidzin, and puerarin) were also performed for the comparison but not used for the following correlation analysis due to lack of experimental observations.

The first 2 ns trajectories for each window were discarded for equilibration, and the remaining was used for data analysis. PMFs, $\Delta G(\xi)$, for β-CD binding reactions were computed by the weighted histogram analysis method (WHAM).[62,63] Statistical errors of $\Delta G(\xi)$ were calculated using the Bayesian bootstrapping of complete histograms.[63] Using the −cyclic option of the "g_wham" tool,[63] all the PMFs were defined to zero at ξ = −2 and 2 nm where host and guest molecules were completely separated and the interactions between the binding partners are negligible. Integrating the PMFs with a cylinder approximation[42,64,65] gives a binding constant (K_a) for β-CD complexes with guest (eq 2)

$$K_a = N_A \int \pi r(\xi)^2 \exp[-\Delta G(\xi)/RT] d\xi \quad (2)$$

$$\Delta G^0_{calc} = -RT \ln(K_a C^0) \quad (3)$$

$$\Delta H^0_{calc} = RT^2 \frac{d}{dT} \ln(K_a C^0) = \frac{\int r(\xi)^2 \Delta G(\xi) \exp[-\Delta G(\xi)/RT] d\xi}{\int r(\xi)^2 \exp[-\Delta G(\xi)/RT] d\xi} \quad (4)$$

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

$$-T\Delta S^0_{calc} = \Delta G^0_{calc} - \Delta H^0_{calc} \quad (5)$$

where N_A is the Avogadro constant, R is the ideal gas constant, $\pi r(\xi)^2$ is the sampled area for the centroid of guest in the X-Y plane at ξ, and C^0 is the standard concentration of 1 mol/L. Given K_a, all standard free energy parameters of binding (ΔG^0_{calc}, ΔH^0_{calc}, and ΔS^0_{calc}) can be calculated readily by eqs 3−5.[37,66] The integration interval (eqs 2−4) ranges from each side of the PMF (−2 for the US mode and 2 for DS) to the first central maximum. If there is no obvious maximum of the PMF, the integral is performed over the whole range of ξ.

2.4. Thermodynamic Decomposition. The binding energy (ΔE_{bind}) of β-CD host−guest complexes is generally defined as a difference in potential energy of the complex state with respect to the separated state of the binding partners (eq 6).

$$\Delta E_{bind} = \Delta E_{complex} - (\Delta E_{host} + \Delta E_{guest}) \quad (6)$$

In the PMF calculations this quantity (denoted as ΔE_{PMF}) is a function of ξ defining the formation process of host−guest complexes. For a detailed understanding of the binding process, ΔE_{PMF} can be decomposed further into seven terms for GB simulations, while for explicit solvent models ΔE_{PMF} constitutes eight components (eqs 7 and 8), due to the difference in the treatment of the solvation contribution (ΔE_{sol}).

$$\Delta E_{PMF} = \Delta E_{host} + \Delta E_{guest} + \Delta E_{host-host} + \Delta E_{guest-guest} + \Delta E_{host-guest} + \Delta E_{sol} \quad (7)$$

$$\Delta E_{sol} = \begin{cases} \Delta E_{pol} + \Delta E_{np} = \Delta E_{pol} + \sigma \cdot SA & \text{for GB solvent} \\ \Delta E_{host-sol} + \Delta E_{guest-sol} + \Delta E_{sol-sol} & \text{for explicit solvent} \end{cases} \quad (8)$$

The first five contributions to ΔE_{PMF} (eq 7) are the ΔE_{MM} part in eq 1 and are identical in these two solvent models, namely, bonded (ΔE_{host}+ΔE_{guest}) and nonbonded ($\Delta E_{host-host}$+$\Delta E_{guest-guest}$) intramolecular interactions of host and guest molecules and the intermolecular interactions of the binding partners ($\Delta E_{host-guest}$). In GB models, the polar solvent contributions (ΔE_{pol}) were calculated with the generalized Born equation,[50,51] and the nonpolar part (ΔE_{np}) was generally obtained from the total solvent-accessible surface area (SA) multiplied by a surface tension (σ).[67] For explicit solvent simulations, ΔE_{sol} is composed of the desolvation of host ($\Delta E_{host-sol}$) and guest ($\Delta E_{guest-sol}$) molecules and of solvent−solvent interactions ($\Delta E_{sol-sol}$) (eq 8). Standard derivations of the quantities in eqs 7 and 8 were estimated from a binning analysis.[68]

Entropy changes (ΔS_{PMF}) upon complexation of the binding partners can be evaluated separately as well, and here we used a quasi-harmonic approximation for this purpose.[69] Similar to ΔE_{PMF}, ΔS_{PMF} is also a function of ξ with respect to the separated state ($\xi \sim -2$ nm) of host and guest molecules. Unlike explicit solvent simulations, solvent molecules are modeled implicitly in GB models, and therefore solvent entropy (ΔS_{sol}) contributes to the binding indirectly only (eq 9).

$$\Delta S_{PMF} = \begin{cases} \Delta S_{host} + \Delta S_{guest} & \text{for GB solvent} \\ \Delta S_{host} + \Delta S_{guest} + \Delta S_{sol} & \text{for explicit solvent} \end{cases} \quad (9)$$

It should be noted here that summing entropy contributions into one ΔS_{PMF} (eq 9) is an approximation in itself. The ΔS components are assumed to have the same error bars as ΔE_{PMF}. The functional dependence on ξ of the quantities in eqs 7−9 is omitted here for simplicity. Further details on the decomposition of energy and entropy in explicit solvent simulations have been presented in ref 37.

All the quantities discussed in this work can be weighted by their Boltzmann factors to achieve a quantitative determination of the energetic using eq 10.

$$\langle \Delta E \rangle = \frac{\sum_i \Delta E_i \exp(-\Delta G_i/RT)}{\sum_i \exp(-\Delta G_i/RT)} \quad (10)$$

When weighting for one binding mode (US/DS), ΔE_i can be ΔE_{PMF}, ΔS_{PMF}, or their components in eqs 7−9, and ΔG_i is the PMF profile (note that all these variables are a function of ξ in this case). If one does the weighting for two binding modes, ΔE_i can be the (free) energy of binding for US and DS modes, and ΔG_i can be the binding free energy of the corresponding binding mode.

2.5. Amber Scoring. Predictions of binding modes and binding energy for the 75 β-CD complexes were done as well based on docking techniques with Amber scoring using the Dock6 software (version 6.7).[21] Note that β-CD and guest parametrizations for PMF simulations (section 2.4) with GROMACS[57−59] and for the docking with Dock6.7[21] follow the same scheme and thus are identical in such two calculations. Following the Dock6 tutorial (http://dock.compbio.ucsf.edu/DOCK_6/tutorials/), the molecular surface area of the receptor β-CD was first determined by the DMS program[70] with a probe radius of 0.14 nm after structure preparation, the spheres surrounding β-CD were generated using the DOCK accessory program sphgen,[71] and then a box was constructed around the spheres plus a 0.5 nm margin in all directions. Prior to docking, the accessory program grid[72] was implemented for precomputation of interactions using energy scoring on a 0.03 nm resolution grid within the box.

The docking was performed with grid-based scoring for flexible ligands (i.e., guests), allowing guest structure rearrangements in response to the receptor. The 10 topmost poses (i.e., those with the highest score) of the resulting 100 searched conformations were then reranked with Amber scoring. The binding energy in the Amber scoring scheme was defined to be the same as that in eq 6. Three available GB implicit models (GB^{HCT} with igb = 1, $GB^{OBC}I$ with igb = 2, and $GB^{OBC}II$ with igb = 5) in Dock6.7[21] were examined in this work. The calculation for receptor, ligand, and the corresponding complex follows the same protocol: 100 cycles of conjugate gradient minimization, followed by MD simulations, and by another 100 cycles of minimization for energy evaluations. For evaluation of docking performance and comparison with the above PMF calculations, the Amber scoring was carried out with varied numbers of GB MD steps, namely 300, 1000, 3000, 10000, or 20000, for a flexible and rigid receptor, respectively. The guest was always flexible in the calculations. The scoring temperature was the same as in the PMF calculations (section 2.4), and the other parameters were set to default values.

3. RESULTS

Potentials of mean force (PMFs) calculations with umbrella sampling and docking with Amber scoring were conducted for 75 β-CD host−guest complexes in order to evaluate the

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

performance of GB models in large scale affinity predictions. In the following section, we compare the calculated binding thermodynamics with available experiments and explicit simulations in order to evaluate the performance of GB/SA scoring for the binding affinity prediction and to discriminate the solvation treatment in explicit and GB implicit solvent models.

3.1. Calculated Binding Energy. Free energy profiles (i.e., PMFs) for the formation process of β-CD complexes with nabumetone in the DS binding mode calculated using GB models are presented in Figure 3. Nabumetone penetrates into

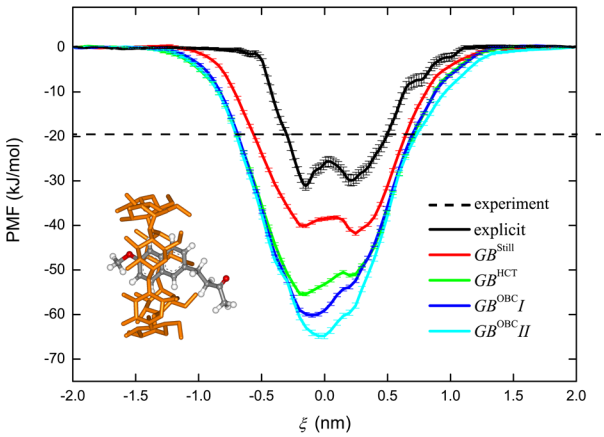


Figure 3. Potential of mean force (PMF) profiles for the [β-CD:nabumetone] complex formation in the DS state with GB^{Still}, GB^{HCT}, $GB^{OBC}I$, and $GB^{OBC}II$ models. The PMF with explicit solvents taken from ref 37 was shifted by 0.2 nm along the $+\xi$ direction and given here for direct comparison with GB models, and an inclusion complex was displayed as well with a stick model. The dashed line indicates the experimental binding free energy of −19.2 kJ/mol taken from ref 73.

the CD hydrophobic cavity from the primary rim along $-Z$, forms an inclusion complex in the DS state, and then moves away from the cavity. The PMFs level off and approach to zero at both ends of ξ, representing completely separated states of the binding partners. A less narrow PMF landscape is observed for GB models compared with the explicit solvent, due to more effective screening by explicit water molecules showing that the interaction between host and guest is effective only at short distances. Form the well depth of the PMFs in Figure 3, it can be seen that GB models overestimate the binding affinity by ~30% for GB^{Still} and by almost 100% for the other three GB models.

PMF profiles with GB^{Still}, GB^{HCT}, $GB^{OBC}I$, and $GB^{OBC}II$ for all the guests studied are given in Figures S2 and S3 for the US and DS binding states, respectively. From these PMFs, standard thermodynamic parameters of binding (ΔG^0_{calc}, ΔH^0_{calc}, and ΔS^0_{calc}) were calculated using the cylinder approximation (eqs 3−5)[37,66] and then weighted for the two binding modes by eq 10. Figure 4 shows a comparison of weighted ΔG^0_{calc} with the corresponding experimental values (ΔG^0_{exp}), revealing a significant discrepancy with low correlation ($R^2 \sim 0.27$) and with large root-mean-square deviations ranging from 8 to 20 kJ/mol, as shown by the absolute RMSD ($RMSD_A$) in Table 1. The calculations appear not to give a reasonable ranking order for the binding guests as indicated by a small Spearman rank order coefficient ($\rho \sim 0.46$). A more detailed comparison is presented in Table 1; ΔG^0_{calc} gives a systematic overestimation

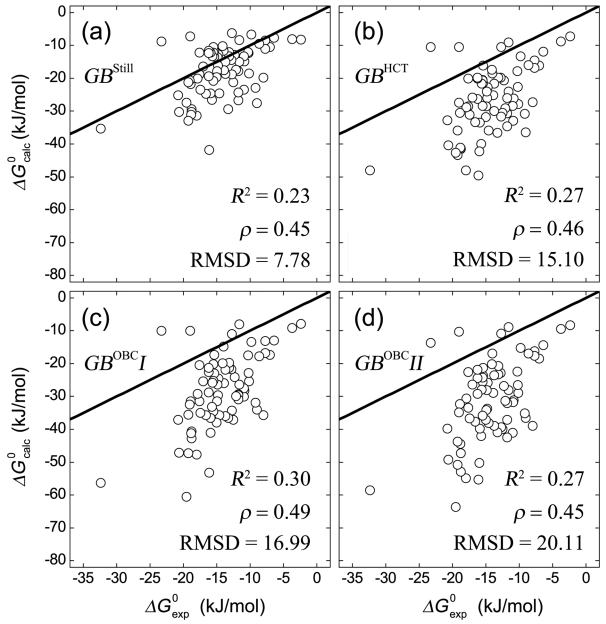


Figure 4. Comparison between experimental and calculated free energy of binding for the 75 [β-CD:guest] complexes using implicit solvent models of (a) GB^{Still}, (b) GB^{HCT}, (c) $GB^{OBC}I$, and (d) $GB^{OBC}II$.

Table 1. Comparison of PMF Calculated Binding Energy with Experimental Observations[a]

	GB^{Still}	GB^{HCT}	$GB^{OBC}I$	$GB^{OBC}II$
$\Delta G^0_{exp} \sim \Delta G^0_{calc}$				
$RMSD_A$	7.8 (0.7)	15.1 (0.9)	17.0(1.0)	20.1 (1.1)
$RMSD_R$	6.8(0.6)	8.6(0.7)	9.5(0.9)	14.7(1.1)
MSE	3.7 (0.8)	12.4 (0.9)	14.1(1.1)	7.0 (1.3)
R^2	0.23 (0.10)	0.27 (0.11)	0.30 (0.12)	0.27 (0.11)
ρ	0.45 (0.12)	0.46 (0.11)	0.49 (0.11)	0.45 (0.11)
slope	0.71 (0.04)	0.51 (0.02)	0.48 (0.02)	0.44 (0.02)
$\Delta G^0_{exp} \sim \Delta E_{PMF}$				
$RMSD_A$	53.0 (1.4)	59.2(1.7)	59.9 (1.7)	63.6(2.0)
$RMSD_R$	12.5(0.9)	13.8(0.9)	14.4(1.0)	16.6(1.4)
MSE	51.6(1.4)	57.5 (1.6)	58.2(1.6)	61.4(1.9)
R^2	0.27 (0.10)	0.27 (0.10)	0.30 (0.10)	0.24 (0.09)
ρ	0.47 (0.11)	0.46 (0.11)	0.49 (0.11)	0.47 (0.11)
slope	0.21 (0.01)	0.20 (0.01)	0.19 (0.01)	0.18 (0.01)

[a]The uncertainties (given in parentheses) of root-mean-square deviation (RMSD, kJ/mol), mean signed error (MSE, kJ/mol), correlation coefficient (R^2), Spearman rank order coefficient (ρ), and the slope for linear fitted lines through the origin were calculated by the bootstrapping with 1000 iterations using the *R* program;[74] $RMSD_A$ stands for an absolute estimation of RMSD and $RMSD_R$ for a relative one after subtraction of MSE.

of ΔG^0_{exp} by 3−14 kJ/mol on average, as indicated by mean signed errors (MSE), and a factor (i.e., the slope of ΔG^0_{exp} vs ΔG^0_{calc}) of 0.4−0.7 can be used to scale ΔG^0_{calc} to fit ΔG^0_{exp}. The GB^{Still} model seems to outperform GB^{HCT} and GB^{OBC} models, as indicated by a smaller $RMSD_A$ and MSE and by a larger scale factor of 0.71 (Table 1). Despite the weak correlation of ΔG^0_{calc} with ΔG^0_{exp}, neither the calculated binding enthalpy (ΔH^0_{calc}) nor the entropy (ΔS^0_{calc}) shows any correlations with experimental determinations (Figures S4 and S5).

Binding energies evaluated by eq 6 are often applied in high-throughput virtual screening tools, including docking techniques, to select efficient inhibitors against targeted receptors.

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

Here we also evaluated the PMF calculated binding energies (ΔE_{PMF}) and made a comparison with the experimental (ΔG^0_{exp}) binding free energies, as shown in Table 1. From the view of correlations with ΔG^0_{exp} and of ρ, ΔE_{PMF} shows the same poor performance as ΔG^0_{calc} but with much larger root-mean-square deviations ($RMSD_A \sim 59$ kJ/mol) from ΔG^0_{exp} and large systematic overestimations (MSE $\sim$ 57 kJ/mol). ΔE_{PMF} can be scaled by a factor of ~0.2 to fit the experimental binding free energies, in close agreement with Levy's report.[25] If the contributions of configurational entropy (ΔS_{PMF}) for the binding partners are considered, the calculated deviations from ΔG^0_{exp} such as RMSD and MSE are reduced to some extent (Table S5). However, such consideration seems not to improve the performances of tested GB models, as revealed by ΔG^0_{exp} vs (ΔE_{PMF}-$T\Delta S_{PMF}$) in Table S5. Relative RMSDs ($RMSD_R$) after subtraction of the MSE are given in Tables 1 and S5 as well, revealing that PMF calculations with the GB models do not predict good relative affinities.

Enthalpy−entropy compensation is acknowledged to be a common phenomenon in cyclodextrin binding with various guests,[33,75] although it is an approximation, that is only correct in theoretical studies, when using pair-potentials.[76] In this work, experimental thermodynamic data (ΔH^0 and $T\Delta S^0$) for β-CD binding with the tested guest molecules show a nearly perfect compensatory enthalpy−entropy relationship (the gray line in Figure 5); that is, enthalpy gain in the binding is canceled out

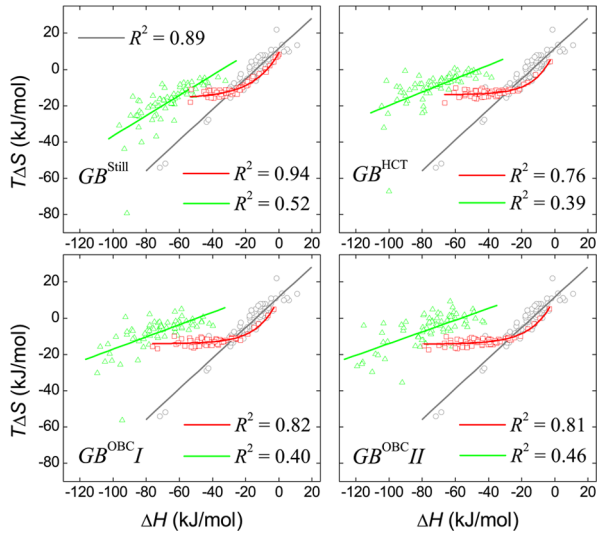


Figure 5. Enthalpy−entropy compensation plot for the β-CD inclusion complexation with the tested guests evaluated by PMF simulations using GB models. Experimental observations (in gray) and calculated binding energies (in green) are shown by a linear fit; the calculated binding free energies (in red) are displayed with fitted exponential curves.

to some extent by entropy loss. In order to evaluate performance of the GB models, the calculated entropy ($T\Delta S$) as a function of enthalpy (ΔH) is plotted and given in Figure 5. Since the box volume and temperature are both controlled in the PMF calculations, the enthalpy change (ΔH) therefore reasonably amounts to the potential energy difference (i.e., ΔE_{PMF}, similar to eq 6) with respect to a completely separated state of the binding partners.[76] The free energy components, binding enthalpy (ΔH^0_{calc}) and binding entropy (ΔS^0_{calc}), calculated with a cylinder approximation (see Section 2.3), are given in Figure 5 (red) as well. ΔS^0_{calc} seems to change very slightly when ΔH^0_{calc} is more negative than ~−30 kJ/mol for the four GB models; with ΔH^0_{calc} larger than this limit, the CD-guest binding appears to follow the rule of enthalpy−entropy compensation, similar to that observed for experimental data of ΔH^0 vs $T\Delta S^0$. An enthalpy−entropy compensation is also found for the PMF calculated ΔH and $T\Delta S$ (green in Figure 5), although the correlation ($R^2 \sim 0.4$−0.5) is not as high as that for the experiments. The discrepancy in enthalpy−entropy relationships between the calculated and experimental observations may result from the omission of solvent-related entropy in the GB implicit treatment.

3.2. Thermodynamic Decomposition. For a deeper understanding of the forces driving the binding reactions, binding energy and configuration entropy are decomposed further into individual contributions from the binding partners and surrounding (explicit/implicit) solvent; for details of the decomposition please refer to Section 2.4 and eqs 7−9. Figure 6 shows the decompositions along the reaction coordinate for

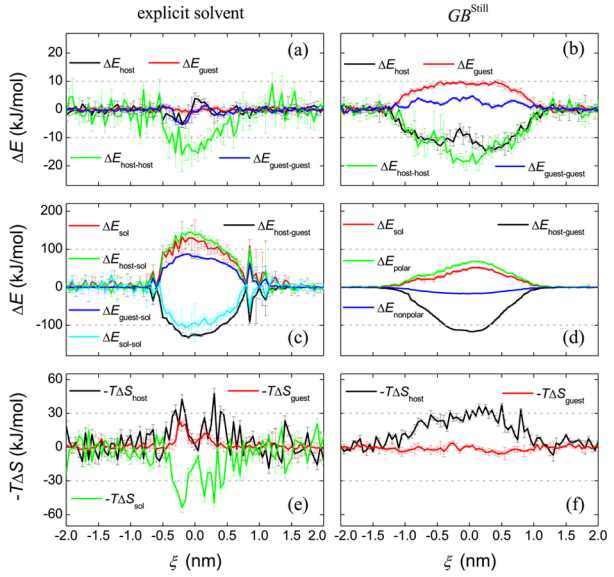


Figure 6. Comparison of energy decomposition between explicit (*left*) and GB^{Still} (*right*) implicit solvent models for the [β-CD:nabumetone] complex formation in the DS state. (a) and (b) represent intramolecular interactions, (c) and (d) for intermolecular interactions, and (e) and (f) for entropy contributions.

[β-CD:nabumetone] DS complex formation with TIP3P explicit and GB^{Still} implicit solvent models. For both models (Figures 6a and 6b), nonbonded intramolecular interactions of the binding partners ($\Delta E_{host-host}$ and $\Delta E_{guest-guest}$) display similar curves; $\Delta E_{host-host}$ favors the binding and $\Delta E_{guest-guest}$ contributes little to the binding. In explicit solvent, bonded intramolecular interactions of host (ΔE_{host}) and guest (ΔE_{guest}) molecules fluctuate around zero contributing little to the complexation (Figure 6a). In the GB^{Still} implicit solvent, however, ΔE_{host} and ΔE_{guest} show clear and opposite contributions for ΔE_{host} and ΔE_{guest}, implying a discrepancy in molecular flexibility between explicit and GB implicit models; the former favors the binding and the latter does not (Figure 6b). These four components mentioned above reflect conformational changes of the binding partners upon complexation due to configuration match (also known as "induced-fit" effect), and adding together these contributions constitutes

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

intramolecular strain energies (ΔE_{strain}) of host and guest molecules. Quantitative evaluations of these four contributions for explicit and GB solvent models are tabulated in Table 2,

Table 2. Decompositions of Binding Energy (kJ/mol) for the [β-CD:Nabumetone] Complex in the DS State Calculated with Explicit and GB Implicit Solvent Models[a]

	explicit[b]	Still	HCT	OBC(*I*)	OBC(*II*)
ΔG^0_{calc}	−18(1)	−30(1)	−44	−48	−51
ΔE_{PMF}	−28(5)	−82(2)	−94	−102	−104
ΔE_{strain}	−13(3)	−15(3)	−9	−12	−12
ΔE_{host}	−1(1)	−11(2)	−10	−11	−14
ΔE_{guest}	0(1)	9(1)	9	10	10
$\Delta E_{host-host}$	−12(3)	−17(2)	−9	−11	−10
$\Delta E_{guest-guest}$	−2(1)	3(1)	2	1	1
$\Delta E_{host-guest}$	−123(8)	−110(2)	−116	−113	−113
ΔE_{sol}	115(15)	43(2)	31	23	21
$-T\Delta S_{host}$	21(3)	27(2)	19	21	14
$-T\Delta S_{guest}$	11(2)	0(2)	−2	−1	0
$-T\Delta S_{sol}$	−33(5)				

[a] ΔE and ΔS were computed with Boltzmann weighting by eq 10. All the GB quantities have similar standard deviations (given in parentheses), and only the deviations of GB^{Still} are shown here. [b] Explicit results taken from ref 37.

indicating that host and guest molecules adjust their configurations to minimize the total energy. Since the opposite contributions of ΔE_{host} and ΔE_{guest} in GB models seem to cancel out (Figure 6b and Table 2), the explicit and GB models ultimately display similar behavior for the intramolecular strain energies (ΔE_{strain}).

Both the explicit and GB models predict the intermolecular interaction between host and guest ($\Delta E_{host-guest}$) within the same order of magnitude, although explicit solvents give a more negative $\Delta E_{host-guest}$ by ∼15 kJ/mol (Figure 6c-d and Table 2). In explicit solvent models, unfavorable host-solvent ($\Delta E_{host-sol}$) and guest-solvent ($\Delta E_{guest-sol}$) interactions (i.e., desolvation effects) and favorable solvent−solvent ($\Delta E_{sol-sol}$) interactions are observed for the [β-CD:nabumetone] complex formation, and these three components constitute a positive solvation contribution (ΔE_{sol}) disfavoring the binding (Figure 6c). The solvation by the GB^{Still} model is decomposed into an unfavorable polar interaction (ΔE_{pol}) and a favorable nonpolar interaction (ΔE_{np}), and the resulting (unfavorable) solvation contribution (ΔE_{sol}) is three times smaller than that of the explicit solvent (Figure 6c-d and Table 2).

Conformations of host and guest molecule are somewhat restricted upon binding, leading to an entropy loss of the binding partners, as indicated by negative ΔS_{host} and ΔS_{guest} in explicit solvents. Meanwhile, explicit solvent molecules display a favorable entropy gain compensating for entropy loss of the binding partners to some extent (Figure 6e). In the GB^{Still} model, the β-CD host shows an entropy loss as well, while the guest entropy seems unchanged when binding to β-CD (Figure 6f). Due to the entropy compensation of solvent molecules, the total entropy contributes little to the binding in explicit solvent model; however, the GB model produces a net entropy loss disfavoring the [β-CD:nabumetone] inclusion formation (Table 2).

Comparison of explicit and GB models for the other four guests (2-propanol, daidzein, daidzin, and puerarin) shows that GB models appear to give favorable entropy contributions (i.e., entropy gain) to the binding (Table S6), as opposite to explicit simulations. The energy decompositions (Figure 6 and Tables 2 and S6) indicate that the discrepancies between explicit and GB solvent models are systematic. Binding predictions with these two models indicate that the explicit solvent by far outperforms all implicit models for the tested guests (Table S7 and Figure S6), although the two solvation models seem to yield similar accuracies for the binding of β-CD with 2-propanol.

For computational efficiency, binding energy (ΔE_{PMF}) is usually used as a ranking score instead of binding free energy (ΔG^0_{calc}) in virtual screening. An analysis on the relationship between these two terms for all the guests and GB models (300 data points in total) is given in Figure S7a showing that ΔE_{PMF} correlates well with ΔG^0_{calc} ($R^2 \sim 0.8$) and a high Spearman rank order coefficient ($\rho \sim 0.9$) is observed. This finding reveals that ΔE_{PMF} reproduces a reasonable ranking order for the binding affinities of guest molecules and justifies the use of ΔE_{PMF} instead of ΔG^0_{calc}. However, it should be kept in mind that the GB models do not accurately predict binding energies (ΔG^0_{exp} vs ΔG^0_{calc} in Table 1). Correlations of ΔG^0_{calc} with the individual components of ΔE_{PMF} indicate that ΔG^0_{calc} has very weak or even no correlations with any component (Table S8). For instance, ΔG^0_{calc} correlates weakly with host−guest interactions ($\Delta E_{host-guest}$, $R^2 \sim 0.4$), and almost no correlation is detected with the solvation contribution (ΔE_{sol}, $R^2 \sim 0$). Interestingly, adding together these two components yields a high correlation with ΔG^0_{calc} ($R^2 \sim 0.8$), as shown in Figure S7b, comparable of ΔE_{PMF} vs ΔG^0_{calc} (Figure S7a), implying that solvation contribution affects the binding significantly and these two contributions need to be considered with care for affinity prediction of ligand binding.

In addition, polar (ΔE_{pol}) and nonpolar (ΔE_{np}) effects are scaled to correlate experimental and calculated binding free energies (Table S8). ΔE_{pol} shows very weak or almost no correlations with experimental and calculated binding free energies, while ΔE_{np} gives a higher correlation than ΔE_{pol}, revealing the nature of β-CD host−guest interactions dominated by hydrophobic interactions. Unlike in GB^{Still} ($R^2 \sim 0.5$), almost no correlations between ΔE_{np} and the binding free energies are observed for two GB^{OBC} models, which explains why the GB^{Still} model performs better than the GB^{OBC} models.

3.3. Binding Prediction with Amber Scoring. Convergence of docking with Amber scoring for binding prediction was first examined with a rigid/flexible β-CD host and with a varied number of GB simulations steps. An examination of the convergence with the GB^{HCT} model reveals (Figure 7) that a rigid β-CD gives a higher correlation ($R^2 \sim 0.84$) and a higher Spearman rank order coefficient ($\rho \sim 0.9$) than that of the flexible host between PMF calculated binding energy (ΔE_{PMF}) and docked binding energy (ΔE_{dock}). Similar to the PMF simulations (Table 1), either rigid or flexible β-CD in the docking experiments gives a weak correlation with ΔG^0_{exp} ($R^2 \sim 0.2$) and a low Spearman rank order coefficient ($\rho \sim 0.4$). Similar observations were made for GB^{OBC} docking calculations with Amber scoring (Figure S8). Moreover, increasing the number of GB simulation steps does not improve the docking performance and in some cases gives a worse prediction of binding affinities yet, as indicated in Figures 7 and S8. If not stated otherwise, the docking results presented below were calculated with rigid β-CD and 3000 GB steps, the default appropriate for protein−ligand docking in Dock 6.7 software.

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

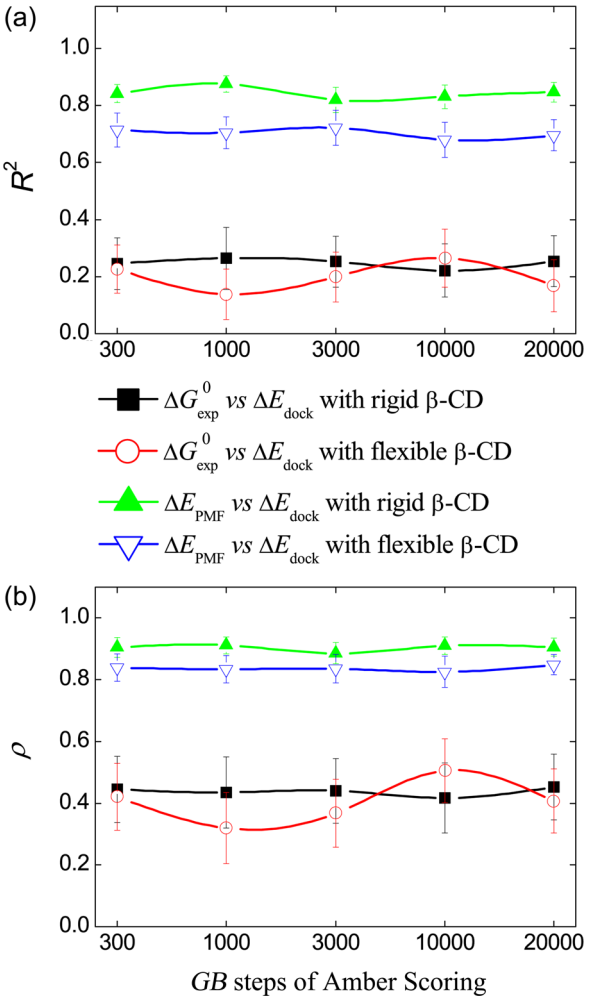


Figure 7. Correlation (R^2) and Spearman rank order (ρ) coefficients between experimental and calculated binding energy as a function of GB^{HCT} simulation steps with Amber scoring for flexible and rigid β-CD.

Table 3 lists the comparison of docked binding energy (ΔE_{dock}) with experimental (ΔG^0_{exp}) and PMF calculated (ΔE_{PMF}) observations. ΔE_{dock} correlates weakly with ΔG^0_{exp} ($R^2 \sim 0.3$) and overestimates it by ~37 kJ/mol, and a factor of 0.27 is needed for scaling ΔE_{dock} to fit to ΔG^0_{exp}. Similar to ΔE_{PMF} (Table 1), ΔE_{dock} does not reproduce a reasonable ranking order for guest binding affinities as well, as revealed by a small ρ of ~0.45. Surprisingly, $\Delta E_{dock,}$ sporting a low computational cost, shows similar performances to the relatively expensive PMF calculations, as indicated by higher R^2 of ~0.7 and ρ of 0.87, even though ΔE_{dock} underestimates ΔE_{PMF} by ~22 kJ/mol. However, neither ΔE_{dock} nor ΔE_{PMF} predict the experimental binding energy accurately. Relative RMSDs ($RMSD_R$) indicate that Amber scoring in the docking do not give good relative affinities (Table 3) compared to the experiments and neither do the PMF calculations (Table 1). Interestingly, however, the $RMSD_R$ between PMF and Amber scoring techniques is less than 1 kJ/mol (Table 3), highlighting the similar performance of these two methods.

3.4. Binding Modes of β-CD Complexes with Guest.

Accuracy of binding mode prediction is also of great importance in biotechnology applications. There are two possible binding modes for β-CD host−guest complexes due to the existence of two different rims on β-CD, as shown in Figure 2. Guest insertion into the CD hydrophobic cavity may be easier from the secondary (wider) rim than from the primary (narrower) rim. The PMF simulations show that approximately 70% of the inclusion complexes are formed via guest penetrating into CD cavity from the secondary rim (i.e., US mode in Figure 2), and the guests bind more stably than in the primary rim (i.e., DS mode in Figure 2). Prediction of guest binding orientations with docking depends on the choice of β-CD flexibility and GB simulation steps. For instance, with a rigid β-CD and 3000 GB steps, about 40% of the guests are more stable in the DS binding mode; a flexible β-CD and 20000 GB steps gives a probability of ~60% favoring the US mode that is comparable to the PMF GB predictions (Table S9), but this choice may decrease the docking performance on ranking the binding affinity (Figures 7 and S8).

Comparison of the PMFs with docking results indicates that only 12 out of 75 [β-CD:guest] complexes are predicted to have similar guest orientations in the binding, including five DS and seven US states. Some representative states of these similar binding orientations are given in Figure 8. The bromo group is typically bound within the β-CD hydrophobic cavity due to its high hydrophobicity (comparable to a methyl group, Table S4), and cycloalkane moieties behave similarly as well. Hydrophobic aromatic rings are sometimes included in the cavity, depending on the molecular geometry and other attached moieties. 63% of the tested binding partners have similar binding modes in all four GB models in PMF calculations. For the three GB models in docking the corresponding number is 61% (Table S10), even though PMF and docking methods favor different binding modes (US/DS) as stated above. Different GB models seem to produce comparable results on predicting guest binding orientations; more specific, approximately 80% of the complexes were shown to have similar guest orientations in the binding prediction but only when using the same method (PMF or docking). Different methods predict different binding modes for about half of the [β-CD:guest] complexes (Table S10).

Table 3. Comparison of Docked Binding Energy with Experimental and PMF Calculated Observations[a]

	GB^{HCT}	$GB^{OBC}I$	$GB^{OBC}II$
$\Delta G^0_{exp} \sim \Delta E_{dock}$			
$RMSD_A$	44.5(1.4)	34.9(1.4)	39.7(2.5)
$RMSD_R$	12.4(1.0)	11.5(0.9)	16.4(2.4)
MSE	42.7 (1.5)	32.9 (1.3)	36.2(1.9)
R^2	0.25 (0.09)	0.33 (0.11)	0.25 (0.11)
ρ	0.44 (0.10)	0.48 (0.11)	0.43 (0.12)
slope	0.25 (0.01)	0.29 (0.01)	0.26 (0.01)
$\Delta E_{PMF} \sim \Delta E_{dock}$			
$RMSD_A$	16.2 (0.8)	26.5 (0.9)	28.4(1.5)
$RMSD_R$	0.82(0.04)	0.76(0.06)	0.56(0.08)
MSE	−14.8(0.8)	−25.3(0.9)	−25.2 (1.5)
R^2	0.82 (0.04)	0.76 (0.05)	0.56 (0.08)
ρ	0.89 (0.03)	0.87 (0.04)	0.84 (0.04)
slope	1.25 (0.01)	1.50 (0.05)	1.41 (0.06)

[a]The uncertainties are given in parentheses and terms for RMSD, MSE, R^2, ρ, and slope are the same as in Table 1. ΔE_{dock} was computed by 3000-step GB simulation with rigid β-CD using Amber scoring.

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

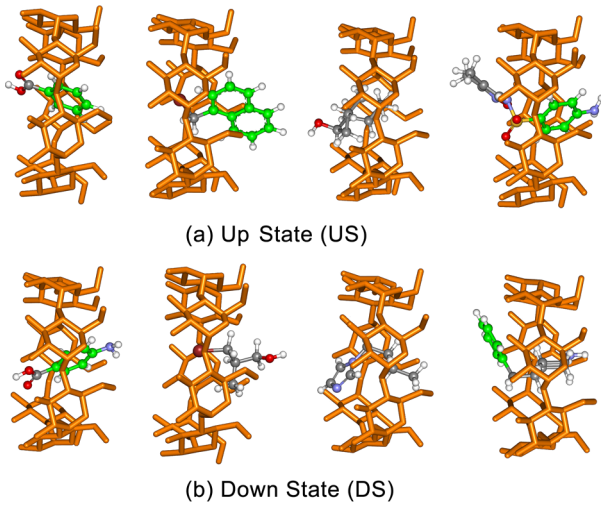


Figure 8. Similar binding orientations of (a) US and (b) DS states predicted by both PMF and dock calculations. From left to right: (a) benzoic acid, 1-bromomethyl naphthalene, cyclooctanol, and sulfisoxazole; (b) 4-aminobenzoic acid, *R*-3-bromo-2-methyl-1-propanol, 1-butylimidazole, and 4-benzylpiperidine. These host−guest complexes are taken from the most stable conformations predicted by GB^{HCT} Amber scoring with rigid β-CD. β-CD is shown by a stick model in orange with its secondary rim pointing left. Guest molecules are displayed with a ball and stick model and colored by element; aromatic rings are colored in green for clarity.

4. DISCUSSION

Binding affinities and binding modes of 75 [β-CD:guest] complexes were investigated through potential of mean force (PMF) calculations with umbrella sampling and docking techniques based upon the Amber scoring using generalized Born (GB) implicit solvent models. Neither of these two methods gives a reasonable prediction of relative binding affinities for this test set compared to the experimental data, as indicated by small Spearman rank order coefficients ($\rho \sim 0.46$) in Tables 1 and 3. Somewhat unexpectedly, the expensive umbrella sampling technique does not appear to improve the binding affinity prediction and shows an accuracy comparable to the docking calculations with Amber GB/SA scoring ($\rho \sim 0.9$). These findings indicate that GB models in the virtual screening model the conformational flexibility reasonably well but that there are other factors that need to be considered.

The free energy of binding ligands to target receptors determines the ranking order for a library of compounds. Due to the high computational cost of free energy calculations, the binding energy (eq 6) is generally used instead in virtual screening procedures. In this work, binding energies derived from the PMF (ΔE_{PMF}) calculations yield a significant overestimation on the experimental binding free energies (ΔG^0_{exp}), and a scaling factor of 0.2 (Table 1) is needed to scale the ΔE_{PMF} to fit with ΔG^0_{exp}, in agreement with a report by Levy.[25] For the docking score, a scaling factor of 0.27 is needed (Table 3). A combination of enthalpy and entropy effects is known to contribute to the binding of protein−ligand complexes,[77−79] and the entropy contribution (ΔS) therefore needs to be taken into account in the binding energy. For the tested [β-CD:guest] complexes, integrating entropy effects into the energy (ΔE_{PMF}-$T\Delta S$) leads to a decrease in the deviation of binding energies from ΔG^0_{exp} by 10 kJ/mol, while such integration does not seem to improve the predictive power for binding affinities. Despite the overestimation, ΔE_{PMF} compares well to the calculated binding free energies (ΔG^0_{calc}), indicated by the correlation with ΔG^0_{exp} (R^2) and Spearman rank order coefficients (ρ) in Table 1, justifying the use of binding energy for scoring in virtual screening. Such use however depends on the accuracy of force field parameters (including solvation treatment discussed below) and entropy calculation algorithms.

Energetic decomposition highlights the difference in solvation treatments between explicit and implicit solvent models and unveils in detail the deficiencies of GB models. Upon complexation, β-CD adjusts its configuration to match its guest molecule (nabumetone), as revealed by favorable nonbonded intramolecular interactions of the host molecule ($\Delta E_{host-host}$), and the guest ($\Delta E_{guest-guest}$) may also contribute to such interactions depending on the kind of guest molecules. Both these interactions behave similarly in explicit and GB implicit solvent models (Figure 6 and Tables 2 and S6). The bonded intramolecular interactions of ΔE_{host} and ΔE_{guest} contribute little to the binding in the explicit solvents; however, significant contributions (favorable ΔE_{host} and unfavorable ΔE_{guest}) are observed in GB solvent models (Tables 2 and S6). Although these two contributions seem to cancel out to some extent, thus displaying a similar and comparable intramolecular strain energy ($\Delta E_{strain} = \Delta E_{host-host} + \Delta E_{guest-guest} + \Delta E_{host} + \Delta E_{guest}$) to that from the explicit solvent, this finding indicates that GB models may generate more flexible configurations (i.e., unreasonable conformational ensembles) than warranted, which may explain the poor accuracy in the docking experiment (Figures 7 and S8) when using a full flexible host and/or increasing GB simulation steps. Hobza and co-workers also indicated that scoring functions based on implicit solvent models may fail for highly flexible receptors.[5] According to the Dock 6.7 tutorial for Amber GB/SA scoring, limited flexibility of receptors (ligand binding sites in general) and use of 3000 GB-steps rather than longer GB simulations are recommended for good performance on the scoring of protein−ligand complexes.

Due to the same parametrization of binding partners in explicit and GB solvent simulations, intermolecular host−guest interactions ($\Delta E_{host-guest}$) are predicted to be of the same order of magnitude, while explicit solvents tend to give a stronger interaction, as shown in Figure 6 and Tables 2 and S6. Desolvation of host ($\Delta E_{host-sol}$) and guest ($\Delta E_{guest-sol}$) molecules are observed to disfavor the binding, and solvent molecules ($\Delta E_{sol-sol}$) offer an obviously favorable contribution to the binding in explicit solvents (Figure 6c). Such solvent-related balance cannot be reached via GB implicit solvent models, and the GB-based solvation contributions (ΔE_{sol}) are shown to be by far smaller (less positive) than those in explicit solvents (Tables 2 and S6). Entropy gain of explicit water molecules cancels out most of the entropy loss of the binding partners upon complexation, while GB models yield a net entropy loss (Tables 2 and S6). It should be noticed that guest molecules in GB models tend to display favorable entropy gains upon complexation, as opposed to explicit simulations and to the general concept that the guest configuration is restricted upon complexation, leading to entropy loss, which is, however, compensated by a gain in water entropy. In addition, β-CD binding with guest molecules is dominated by hydrophobic interactions, which are solvent-induced and in this case likely dominated by the entropy of the solvent. Modeling of solvent entropy is challenging for implicit solvent calculations for obvious reasons. Due to the lack of solvent balance, in

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

particular in gains of enthalpy and entropy, the GB models yield systematic overestimations for the binding affinities of the [β-CD:guest] complexes (Figure 2 and Table 1). These observations highlight detailed differences in the solvation treatment between explicit and implicit solvent models and indicate that more properties such as solvent enthalpy and solvent entropy need to be incorporated into the procedure of developing GB solvation models.

In order to derive an implicit solvent model, in principle, the mean force $\langle F_{sol} \rangle$ exerted by the external media on the biomolecules is averaged overall and applied to the solute. If one just considers the potential energy (V), this does not matter, since we have $V_{total} = V_{slt\text{-}slt} + V_{slt\text{-}sol} + V_{sol\text{-}sol}$ and $dV_{total}/d\boldsymbol{r} = dV_{slt\text{-}slt}/d\boldsymbol{r} + dV_{slt\text{-}sol}/d\boldsymbol{r}$, where slt and sol stand for solute and solvent, respectively, and $\boldsymbol{r}$ is the position vector of the solute. For $dV_{slt\text{-}sol}/d\boldsymbol{r}$, $\langle F_{sol} \rangle$ is substituted. Since the solvent is implicit, solvent-related reactions are ignored, and consequently we lose detailed balance in MD simulations. Lack of detailed balance means that for a molecule in GB/SA simulations without temperature coupling, the energy will not be conserved. Energy drifts in microcanonical-ensemble (NVE) simulations have indeed been reported for GB implementations in the GROMACS,[50] AMBER,[80] and CHARMM[81] simulation toolkits. Also, the energy function becomes a mixture of enthalpy and free energy, which means that the forces derived from such an energy function are not true forces, and hence the concept of "time" in the simulation becomes meaningless (although this is not relevant for computations of PMF, it may be for other applications). It is well established that due to the lack of solvent viscosity in implicit solvent models, folding of proteins in implicit solvent simulations proceeds "faster" than in explicit solvents,[82−85] something that can be "corrected" for by slowing down the simulation using stochastic kicks, corresponding to the solvent motion.

Interactions with solvent (water) molecules carry valuable information for the analysis of the energetics as well as the structure and dynamics of biomolecules,[86−89] and some of this information needs to be preserved reasonably when attempting simplifications to solvation effects. However, some properties of real solvent molecules such as the ability of hydrogen bonding donating/accepting and dipole directionality are difficult to describe in analytical formulas.[90] Some strategies have been proposed to improve the binding energy calculations in either implicit or explicit manners. Gallicchio and co-workers developed an implicit solvent model AGBNP2 (Analytical Generalized Born plus NonPolar), in which a hydration correction term was introduced into the typical GB scheme accounting for, for instance, hydrogen bonding with solvent and water ordering in the binding sites.[91,92] By including two hydration sites of β-CD with empirical parameters further, they subsequently reproduced binding free energies of β-CD with guest molecules reasonably compared to the experiment.[25] Explicit water molecules were also attempted as part of the binding site to improve the performance of implicit solvent models;[93−96] it is however not trivial to position these water molecules "correctly".[97]

In addition, the problem in evaluating scoring function for prediction of binding affinity depends on the reliability of the experimental data and on the techniques used to determine the data. As stated by Bonvin et al., significant correlations between experimental and calculated dissociation constants may be found if the protein−protein complexes are grouped according to the methods employed in the experimental determinations.[98] Similar treatments were made for the tested complexes in this work, indicating that the poor correlations (Tables 1 and 3) between experimental and calculated binding energies with all the GB models are statistically significant for all the β-CD complexes (N = 75, p-vale < 0.01, Table S11). The calorimetry method (N = 39, p-vale < 0.01) gives slightly higher correlations than all the binding affinity data. Significant correlations emerge (p-vale < 0.05), with an R approaching 0.95 for the spectrophotometry (N = 5) and potentiometry (N = 4) techniques, although these results need to be treated with care due to the limited amount of data.

From this and our previous work on GB implicit solvent simulations of cyclodextrin dimerization in nonaqueous phases,[30] we conclude that the solvation enthalpy and entropy as well as the flexibility of the binding partners need to be taken into consideration with caution since these quantities are within the same order of magnitude or even an order of magnitude larger than thermodynamics parameters of the entire binding reactions in both explicit and implicit solvent models.[37,99] Lack of detailed balance in implicit solvent models may result in problematic issues such as poor energy conservation and desolvation of the binding partners. Here we note that solvent screening using explicit models is much larger than with GB models (Figure 2). Dipole analysis shows that explicit water interacts with β-CD, reducing its dipole (Figure S9) compared to GB models and thereby reducing the interaction energy. These findings indicate that the screening of the Coulomb interactions in GB models underestimates the screening due to solvent but that it overestimates the intermolecular (binding) energy (by favoring short-range solute−solute interactions), leading to largely overestimated binding energy. Systematic errors caused by the unphysical treatment of Coulomb interactions seem to be difficult to compensate for. It is difficult to envision a way in which these and other drawbacks of implicit models can be overcome systematically.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.6b00418.

Molecular structure of guest molecules, GB parameters, and supplementary tables and figures (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*Phone: 86 10 6233317. E-mail: haiyan@ustb.edu.cn (H.Y.).
*Phone: 46 18 471 4205. E-mail: david.vanderspoel@icm.uu.se (D.v.d.S).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Prof. Tianwei Tan for a grant of computer time through ChemCloudComputing of Beijing University of Chemical Technology. This work was supported by the National Natural Science Foundation of China (21677011, 21606016, 11471034, and 21306006), China Postdoctoral Science Foundation (2015M580993), Beijing Municipal Science and Technology Commission (z131102002813058), and the Fundamental Research Funds for the Central Universities (FRF-TP-15-018A1) as well as by the Swedish Research Council (grant 2013-5947).

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

REFERENCES

(1) Bajorath, J. Integration of Virtual and High-Throughput Screening. *Nat. Rev. Drug Discovery* **2002**, *1*, 882−894.

(2) Zamora, W.; Campanera, J. M.; Luque, F. Implicit Solvation Methods in the Study of Ligand−Protein Interactions. In *In Silico Drug Discovery and Design: Theory, Methods, Challenges, and Applications*; Cavasotto, C. N., Ed.; CRC Press: Boca Raton, 2015; pp 249−274.

(3) Zhou, R.; Berne, B. J. Can a Continuum Solvent Model Reproduce the Free Energy Landscape of a β-Hairpin Folding in Water? *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12777−12782.

(4) Lin, J.-H.; Baker, N. A.; McCammon, J. A. Bridging Implicit and Explicit Solvent Approaches for Membrane Electrostatics. *Biophys. J.* **2002**, *83*, 1374−1379.

(5) Kolář, M.; Fanfrlík, J.; Lepšík, M.; Forti, F.; Luque, F. J.; Hobza, P. Assessing the Accuracy and Performance of Implicit Solvent Models for Drug Molecules: Conformational Ensemble Approaches. *J. Phys. Chem. B* **2013**, *117*, 5950−5962.

(6) Kleinjung, J.; Fraternali, F. Design and Application of Implicit Solvent Models in Biomolecular Simulations. *Curr. Opin. Struct. Biol.* **2014**, *25*, 126−134.

(7) Baker, N. A. Improving Implicit Solvent Simulations: A Poisson-Centric View. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137−143.

(8) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 383−394.

(9) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.* **2007**, *3*, 156−169.

(10) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. Performance Comparison of Generalized Born and Poisson Methods in the Calculation of Electrostatic Solvation Energies for Protein Structures. *J. Comput. Chem.* **2004**, *25*, 265−284.

(11) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2020−2034.

(12) Kongsted, J.; Söderhjelm, P.; Ryde, U. How Accurate Are Continuum Solvation Models for Drug-Like Molecules? *J. Comput.-Aided Mol. Des.* **2009**, *23*, 395−409.

(13) Mikulskis, P.; Genheden, S.; Rydberg, P.; Sandberg, L.; Olsen, L.; Ryde, U. Binding Affinities in the SAMPL3 Trypsin and Host−Guest Blind Tests Estimated with the MM/PBSA and LIE Methods. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 527−541.

(14) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-Based Scoring of Protein−Ligand Complexes: Enrichment of Known Inhibitors in Large-Scale Virtual Screening. *J. Chem. Inf. Model.* **2006**, *46*, 243−253.

(15) Greenidge, P. A.; Kramer, C.; Mozziconacci, J. C.; Sherman, W. Improving Docking Results Via Reranking of Ensembles of Ligand Poses in Multiple X-Ray Protein Conformations with MM-GBSA. *J. Chem. Inf. Model.* **2014**, *54*, 2697−2717.

(16) Zhang, X.; Wong, S. E.; Lightstone, F. C. Toward Fully Automated High Performance Computing Drug Discovery: A Massively Parallel Virtual Screening Pipeline for Docking and Molecular Mechanics/Generalized Born Surface Area Rescoring to Improve Enrichment. *J. Chem. Inf. Model.* **2014**, *54*, 324−337.

(17) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. Dock 6: Combining Techniques to Model RNA−Small Molecule Complexes. *RNA* **2009**, *15*, 1219−1230.

(18) Qi, J.; Dong, Z.; Liu, J.; Peery, R. C.; Zhang, S.; Liu, J.-Y.; Zhang, J.-T. Effective Targeting of the Survivin Dimerization Interface with Small-Molecule Inhibitors. *Cancer Res.* **2016**, *76*, 453−462.

(19) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449−461.

(20) Massova, I.; Kollman, P. A. Combined Molecular Mechanical and Continuum Solvent Approach (MM-PBSA/GBSA) to Predict Ligand Binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.

(21) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. Dock 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **2015**, *36*, 1132−1156.

(22) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing Scoring Functions for Protein−Ligand Interactions. *J. Med. Chem.* **2004**, *47*, 3032−3047.

(23) Genheden, S. Mm/Gbsa and LLE Estimates of Host−Guest Affinities: Dependence on Charges and Solvation Model. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 1085−1093.

(24) Sun, H.; Li, Y.; Shen, M.; Tian, S.; Xu, L.; Pan, P.; Guan, Y.; Hou, T. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 5. Improved Docking Performance Using High Solute Dielectric Constant MM/GBSA and MM/PBSA Rescoring. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22035−22045.

(25) Wickstrom, L.; He, P.; Gallicchio, E.; Levy, R. M. 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*, 3136−3150.

(26) Kolář, M.; Fanfrlík, J. i.; Hobza, P. Ligand Conformational and Solvation/Desolvation Free Energy in Protein−Ligand Complex Formation. *J. Phys. Chem. B* **2011**, *115*, 4718−4724.

(27) Yu, G.; Jie, K.; Huang, F. Supramolecular Amphiphiles Based on Host−Guest Molecular Recognition Motifs. *Chem. Rev.* **2015**, *115*, 7240−7303.

(28) Zhang, H.; Tan, T.; Feng, W.; van der Spoel, D. Molecular Recognition in Different Environments: Beta-Cyclodextrin Dimer Formation in Organic Solvents. *J. Phys. Chem. B* **2012**, *116*, 12684−12693.

(29) Steffen, A.; Thiele, C.; Tietze, S.; Strassnig, C.; Kämper, A.; Lengauer, T.; Wenz, G.; Apostolakis, J. Improved Cyclodextrin−Based Receptors for Camptothecin by Inverse Virtual Screening. *Chem. - Eur. J.* **2007**, *13*, 6801−6809.

(30) Zhang, H.; Tan, T.; Van der Spoel, D. Generalized Born and Explicit Solvent Models for Free Energy Calculations in Organic Solvents: Cyclodextrin Dimerization. *J. Chem. Theory Comput.* **2015**, *11*, 5103−5113.

(31) Rekharsky, M. V.; Inoue, Y. Complexation Thermodynamics of Cyclodextrins. *Chem. Rev.* **1998**, *98*, 1875−1917.

(32) Rekharsky, M. V.; Goldberg, R. N.; Schwarz, F. P.; Tewari, Y. B.; Ross, P. D.; Yamashoji, Y.; Inoue, Y. Thermodynamic and Nuclear Magnetic Resonance Study of the Interactions of α- and β-Cyclodextrin with Model Substances: Phenethylamine, Ephedrines, and Related Substances. *J. Am. Chem. Soc.* **1995**, *117*, 8830−8840.

(33) Rekharsky, M.; Inoue, Y. Chiral Recognition Thermodynamics of β-Cyclodextrin:The Thermodynamic Origin of Enantioselectivity and the Enthalpy−Entropy Compensation Effect. *J. Am. Chem. Soc.* **2000**, *122*, 4418−4435.

(34) Rekharsky, M. V.; Mayhew, M. P.; Goldberg, R. N.; Ross, P. D.; Yamashoji, Y.; Inoue, Y. Thermodynamic and Nuclear Magnetic Resonance Study of the Reactions of α- and β-Cyclodextrin with Acids, Aliphatic Amines, and Cyclic Alcohols. *J. Phys. Chem. B* **1997**, *101*, 87−100.

(35) Rekharsky, M. V.; Inoue, Y. Solvent and Guest Isotope Effects on Complexation Thermodynamics of α-, β-, and 6-Amino-6-Deoxy-β-Cyclodextrins. *J. Am. Chem. Soc.* **2002**, *124*, 12361−12371.

(36) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(37) Zhang, H.; Tan, T.; Hetényi, C.; van der Spoel, D. Quantification of Solvent Contribution to the Stability of Noncovalent Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 4542−4551.

(38) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(39) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

(40) Chen, W.; Chang, C.-E.; Gilson, M. K. Calculation of Cyclodextrin Binding Affinities: Energy, Entropy, and Implications for Drug Design. *Biophys. J.* **2004**, *87*, 3035−3049.
(41) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. Drugbank: A Comprehensive Resource for in Silico Drug Discovery and Exploration. *Nucleic Acids Res.* **2006**, *34*, D668−D672.
(42) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. Pubchem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44*, D1202−D1213.
(43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G.; Scuseria, E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*; Gaussian, Inc.: Wallingford, CT, 2004.
(44) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self−Consistent Molecular Orbital Methods. Xii. Further Extensions of Gaussian−Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257−2261.
(45) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.
(46) van der Spoel, D.; van Maaren, P. J.; Caleman, C. GROMACS Molecule & Liquid Database. *Bioinformatics* **2012**, *28*, 752−753.
(47) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61−74.
(48) Sharff, A. J.; Rodseth, L. E.; Quiocho, F. A. Refined 1.8-Angstrom Structure Reveals the Mode of Binding of Beta-Cyclodextrin to the Maltodextrin Binding-Protein. *Biochemistry* **1993**, *32*, 10553−10559.
(49) Cezard, C.; Trivelli, X.; Aubry, F.; Djedaini-Pilard, F.; Dupradeau, F. Y. Molecular Dynamics Studies of Native and Substituted Cyclodextrins in Different Media: 1. Charge Derivation and Force Field Performances. *Phys. Chem. Chem. Phys.* **2011**, *13*, 15103−15121.
(50) Larsson, P.; Lindahl, E. A High-Performance Parallel-Generalized Born Implementation Enabled by Tabulated Interaction Rescaling. *J. Comput. Chem.* **2010**, *31*, 2593−2600.
(51) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(52) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.* **1996**, *100*, 19824−19839.
(53) Ponder, J. W.; Richards, F. M. An Efficient Newton-Like Method for Molecular Mechanics Energy Minimization of Large Molecules. *J. Comput. Chem.* **1987**, *8*, 1016−1024.
(54) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AmberTools* 15; University of California: San Francisco, 2015.
(55) Mobley, D. L.; Dill, K. A.; Chodera, J. D. Treating Entropy and Conformational Changes in Implicit Solvent Simulations of Small Molecules. *J. Phys. Chem. B* **2008**, *112*, 938−946.
(56) Mobley, D. L.; Guthrie, J. P. Freesolv: A Database of Experimental and Calculated Hydration Free Energies, with Input Files. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 711−720.
(57) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(58) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(59) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(60) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(61) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. Lincs: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(62) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(63) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g_wham−a Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713−3720.
(64) Auletta, T.; de Jong, M. R.; Mulder, A.; van Veggel, F. C. J. M.; Huskens, J.; Reinhoudt, D. N.; Zou, S.; Zapotoczny, S.; Schönherr, H.; Vancso, G. J.; Kuipers, L. Beta-Cyclodextrin Host−Guest Complexes Probed under Thermodynamic Equilibrium: Thermodynamics and Afm Force Spectroscopy. *J. Am. Chem. Soc.* **2004**, *126*, 1577−1584.
(65) Yu, Y.; Chipot, C.; Cai, W.; Shao, X. Molecular Dynamics Study of the Inclusion of Cholesterol into Cyclodextrins. *J. Phys. Chem. B* **2006**, *110*, 6372−6378.
(66) Filippini, G.; Goujon, F.; Bonal, C.; Malfreyt, P. Energetic Competition Effects on Thermodynamic Properties of Association between Beta-CD and Fc Group: A Potential of Mean Force Approach. *J. Phys. Chem. C* **2012**, *116*, 22350−22358.
(67) Schaefer, M.; Bartels, C.; Karplus, M. Solution Conformations and Thermodynamics of Structured Peptides: Molecular Dynamics Simulation with an Implicit Solvation Model. *J. Mol. Biol.* **1998**, *284*, 835−848.
(68) Hess, B. Determining the Shear Viscosity of Model Liquids from Molecular Dynamics Simulations. *J. Chem. Phys.* **2002**, *116*, 209−217.
(69) Andricioaei, I.; Karplus, M. On the Calculation of Entropy from Covariance Matrices of the Atomic Fluctuations. *J. Chem. Phys.* **2001**, *115*, 6289−6292.
(70) DMS. http://www.cgl.ucsf.edu/Overview/software.html (accessed Sept 21, 2016).
(71) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure. *J. Med. Chem.* **1988**, *31*, 722−729.
(72) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking with Grid-Based Energy Evaluation. *J. Comput. Chem.* **1992**, *13*, 505−524.
(73) Valero, M.; Costa, S. M. B.; Ascenso, J. R.; Mercedes Velázquez, M.; Rodríguez, L. J. Complexation of the Non-Steroidal Anti-

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092

Inflammatory Drug Nabumetone with Modified and Unmodified Cyclodextrins. *J. Inclusion Phenom. Mol. Recognit. Chem.* **1999**, *35*, 663−677.
(74) R Core Team. *R: A Language and Environment for Statistical Computing, version 3.2.3*; R Foundation for Statistical Computing: Vienna, Austria, 2015.
(75) Al Omari, M. M.; Zughul, M. B.; Davies, J. E. D.; Badwan, A. A. Thermodynamic Enthalpy−Entropy Compensation Effects Observed in the Complexation of Basic Drug Substrates with B-Cyclodextrin. *J. Inclusion Phenom. Mol. Recognit. Chem.* **2007**, *57*, 379−384.
(76) Hub, J. S.; Caleman, C.; van der Spoel, D. Organic Molecules on the Surface of Water Droplets - an Energetic Perspective. *Phys. Chem. Chem. Phys.* **2012**, *14*, 9537−9545.
(77) Stahl, M.; Rarey, M. Detailed Analysis of Scoring Functions for Virtual Screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.
(78) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−949.
(79) Gallicchio, E.; Deng, N.; He, P.; Wickstrom, L.; Perryman, A. L.; Santiago, D. N.; Forli, S.; Olson, A. J.; Levy, R. M. Virtual Screening of Integrase Inhibitors by Large Scale Binding Free Energy Calculations: The SAMPL4 Challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 475−490.
(80) Tsui, V.; Case, D. A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers* **2000**, *56*, 275−291.
(81) Chocholoušová, J.; Feig, M. Balancing an Accurate Representation of the Molecular Surface in Generalized Born Formalisms with Integrator Stability in Molecular Dynamics Simulations. *J. Comput. Chem.* **2006**, *27*, 719−729.
(82) Zagrovic, B.; Pande, V. Solvent Viscosity Dependence of the Folding Rate of a Small Protein: Distributed Computing Study. *J. Comput. Chem.* **2003**, *24*, 1432−1436.
(83) Lei, H.; Duan, Y. Two-Stage Folding of HP-35 from *ab initio* Simulations. *J. Mol. Biol.* **2007**, *370*, 196−206.
(84) Onufriev, A. Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview. In *Annual Reports in Computational Chemistry*; Ralph, A. W., David, C. S., Eds.; Elsevier: Amsterdam, 2008; Vol. *4*, pp 125−137.
(85) Feig, M. Kinetics from Implicit Solvent Simulations of Biomolecules as a Function of Viscosity. *J. Chem. Theory Comput.* **2007**, *3*, 1734−1748.
(86) De Simone, A.; Dodson, G. G.; Verma, C. S.; Zagari, A.; Fraternali, F. Prion and Water: Tight and Dynamical Hydration Sites Have a Key Role in Structural Stability. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7535−7540.
(87) Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016**, *116*, 7673−7697.
(88) Fornili, A.; Autore, F.; Chakroun, N.; Martinez, P.; Fraternali, F. Protein−Water Interactions in MD Simulations: POPS/POPSCOMP Solvent Accessibility Analysis, Solvation Forces and Hydration Sites. In *Computational Drug Discovery and Design*; Baron, R., Ed.; Springer New York: New York, NY, 2012; pp 375−392.
(89) de Ruyck, J.; Brysbaert, G.; Blossey, R.; Lensink, M. F. Molecular Docking as a Popular Tool in Drug Design, an in Silico Travel. *Adv. Appl. Bioinform. Chem.* **2016**, *9*, 1−11.
(90) Hub, J. S.; Wolf, M. G.; Caleman, C.; van Maaren, P. J.; Groenhof, G.; van der Spoel, D. Thermodynamics of Hydronium and Hydroxide Surface Solvation. *Chem. Sci.* **2014**, *5*, 1745−1749.
(91) Gallicchio, E.; Levy, R. M. Agbnp: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High−Resolution Modeling. *J. Comput. Chem.* **2004**, *25*, 479−499.
(92) Gallicchio, E.; Paris, K.; Levy, R. M. The AGBNP2 Implicit Solvation Model. *J. Chem. Theory Comput.* **2009**, *5*, 2544−2564.
(93) Yu, Z.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A. First-Shell Solvation of Ion Pairs: Correction of Systematic Errors in Implicit Solvent Models. *J. Phys. Chem. B* **2004**, *108*, 6643−6654.
(94) Mikulskis, P.; Genheden, S.; Ryde, U. Effect of Explicit Water Molecules on Ligand-Binding Affinities Calculated with the MM/GBSA Approach. *J. Mol. Model.* **2014**, *20*, 1−11.
(95) Li, L.; Xu, W.; Lü, Q. Improving Protein-Ligand Docking with Flexible Interfacial Water Molecules Using Swrosettaligand. *J. Mol. Model.* **2015**, *21*, 1−13.
(96) van Dijk, A. D. J.; Bonvin, A. M. J. J. Solvated Docking: Introducing Water into the Modelling of Biomolecular Complexes. *Bioinformatics* **2006**, *22*, 2340−2347.
(97) Jeszenői, N.; Horváth, I.; Bálint, M.; van der Spoel, D.; Hetényi, C. Mobility-Based Prediction of Hydration Structures of Protein Surfaces. *Bioinformatics* **2015**, *31*, 1959−1965.
(98) Kastritis, P. L.; Bonvin, A. M. J. J. Are Scoring Functions in Protein−Protein Docking Ready to Predict Interactomes? Clues from a Novel Binding Affinity Benchmark. *J. Proteome Res.* **2010**, *9*, 2216−2225.
(99) Zhang, H.; Tan, T.; Hetényi, C.; Lv, Y.; van der Spoel, D. Cooperative Binding of Cyclodextrin Dimers to Isoflavone Analogues Elucidated by Free Energy Calculations. *J. Phys. Chem. C* **2014**, *118*, 7163−7173.

DOI: 10.1021/acs.jcim.6b00418
J. Chem. Inf. Model. 2016, 56, 2080−2092