

# Blind prediction of SAMPL4 cucurbit[7]uril binding affinities with the mining minima method

Hari S. Muddana · Jian Yin · Neil V. Sapra ·  
Andrew T. Fenley · Michael K. Gilson

Received: 16 November 2013 / Accepted: 31 January 2014 / Published online: 8 February 2014  
© Springer International Publishing Switzerland 2014

**Abstract** Accurate methods for predicting protein–ligand binding affinities are of central interest to computer-aided drug design for hit identification and lead optimization. Here, we used the mining minima (M2) method to predict cucurbit[7]uril binding affinities from the SAMPL4 blind prediction challenge. We tested two different energy models, an empirical classical force field, CHARMM with VCharge charges, and the Poisson–Boltzmann surface area solvation model; and a semiempirical quantum mechanical (QM) Hamiltonian, PM6-DH+, coupled with the COSMO solvation model and a surface area term for nonpolar solvation free energy. Binding affinities based on the classical force field correlated strongly with the experiments with a correlation coefficient ( $R^2$ ) of 0.74. On the other hand, binding affinities based on the QM energy model correlated poorly with experiments ( $R^2 = 0.24$ ), due largely to two major outliers. As we used extensive conformational search methods, these results point to possible inaccuracies in the PM6-DH+ energy model or the COSMO solvation model. Furthermore, the different binding free energy components, solute energy, solvation free energy, and configurational entropy showed significant deviations between the classical M2 and quantum M2 calculations. Comparison of different classical M2 free energy components to experiments show that the change in the total energy, i.e. the solute energy plus the solvation free energy, is the key driving force for binding, with a reasonable correlation to experiment

( $R^2 = 0.56$ ); however, accounting for configurational entropy further improves the correlation.

**Keywords** SAMPL4 · Supramolecular · Binding affinity · Host–guest · Force field · Semiempirical quantum

## Introduction

Accurate methods for predicting protein–ligand binding affinities would greatly enhance the drug discovery process by accelerating progress and reducing costs [1, 2]. While relatively simple approaches, such as docking and ligand-similarity [3], are routinely used to discover initial hits through virtual screening of compound libraries, more sophisticated free energy calculations should afford greater precision across subtle changes in ligand chemical structures, and thus promise to be particularly useful for lead optimization [4]. However, evidence for accurate prediction of binding affinities using free energy calculations is still largely anecdotal. Aspects of these calculations that merit attention include the accuracy of the energy models (force fields) and the adequacy of conformational sampling and convergence.

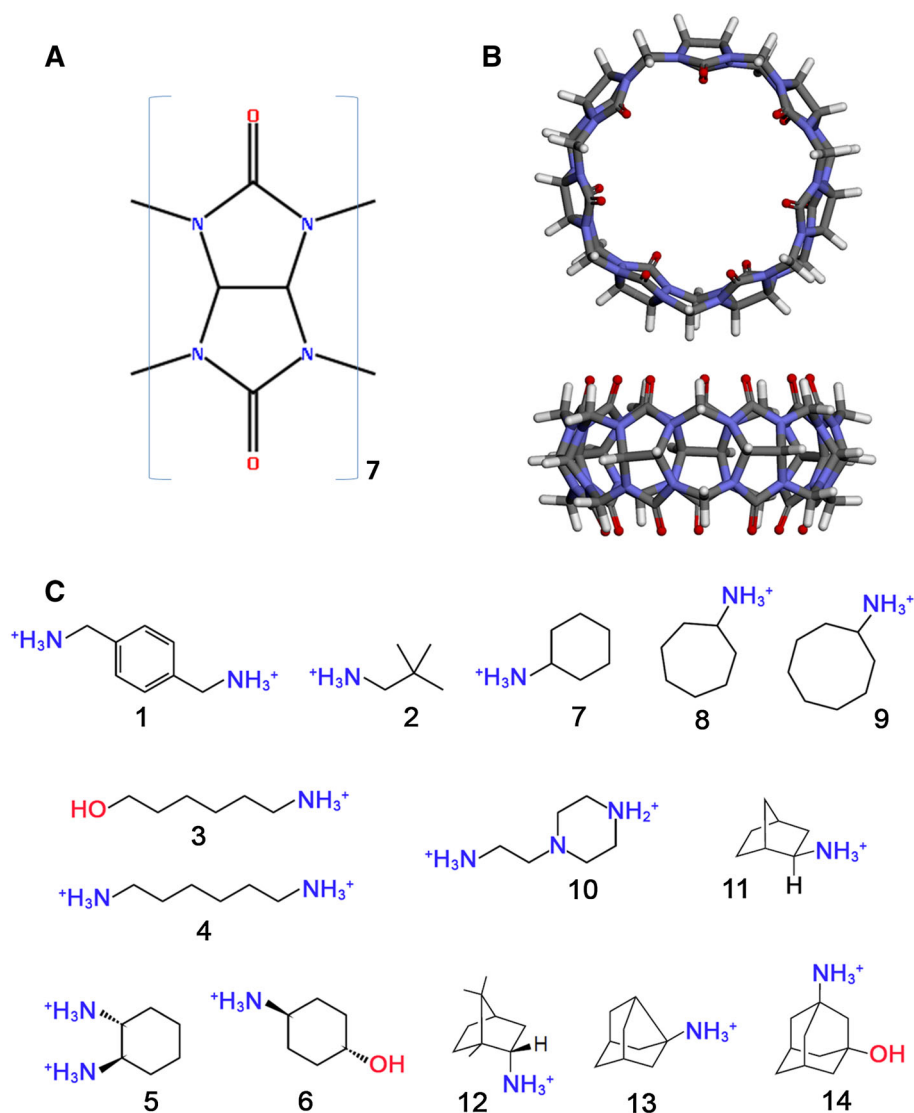
As methods evolve to address these issues, it is important that we critically assess our progress and identify problem areas that require further improvement. The statistical assessment of the modeling of proteins and ligands (SAMPL) blind prediction challenge provides a valuable platform for comparing different methods to each other through validation against experimental data [5–9]. The prospective nature of the challenge reduces bias and provides for a realistic outlook on the performance of the different methods and the current state of the field.

Although it is necessary to test computational methods for biomolecular systems, as done in the SAMPL4 protein–

---

H. S. Muddana · J. Yin · N. V. Sapra · A. T. Fenley ·  
M. K. Gilson (✉)  
Skaggs School of Pharmacy and Pharmaceutical Sciences,  
University of California San Diego, 9500 Gilman Drive,  
Room# 3224, La Jolla, CA 92093-0736, USA  
e-mail: mgilson@ucsd.edu

**Fig. 1** **a** Chemical structure of glycoluril, the repeating unit of CB7. **b** Crystallographic structure of CB7 shown in *top* and *side* views. **c** Chemical structures of SAMPL4 CB7 guest molecules. Protonation and tautomer states of the guest molecules are shown as used in the binding calculations



ligand binding challenge [10], free energy calculations of protein–ligand systems can be problematic, due to their many degrees of freedom (often several thousand) and their conformational flexibility [11]. These make it difficult to assure numerical convergence, and hence to distinguish between convergence problems and force field errors when comparing computational results with experiment. In contrast, supramolecular host–guest systems are computationally quite tractable, as they typically comprise only a few hundred atoms and tend to be more rigid than proteins. As a consequence, convergence is relatively easy to achieve [12], so comparisons with experiment can report clearly on the accuracy of the energy models used [8], and the growing array of host–guest systems today provides many opportunities to test computational models of binding. In addition, host–guest systems are interesting in their own right, due to their practical applications, such as chemical sensing and drug delivery [13, 14], and their

convenience as model systems whose study can deepen our understanding of the physical chemistry of molecular recognition.

Here, we test the mining minima (M2) method in predictions of the binding affinities of cucurbit[7]uril (CB7) with guest molecules from the SAMPL4 challenge (Fig. 1) [15]. The M2 technique combines either an empirical force field or a fast electronic structure method, with an implicit solvent model, and uses conformational search and the harmonic approximation (including, when possible, the mode-scanning correction for anharmonicity) to evaluate binding free energies that incorporate both conformational relaxation and the changes in configurational entropy of the host and guest. In the past, the M2 method using a classical empirical force field was applied successfully in both retrospective and prospective studies [16, 17]. Nonetheless, our predictions for the previous SAMPL3 challenge using M2 deviated significantly from the experimental

measurements [18], and we suspected that the errors likely originated from the underlying classical force field. We therefore developed a new approach, which combined the M2 method with the PM6-DH+ semiempirical quantum mechanical (QM) Hamiltonian and the COSMO solvation model [19]. The resulting QM/M2 method showed promise in initial tests for various CB7 systems [19]. Here, we report new results from M2 calculations using both the classical and semiempirical QM approaches.

## Methods

### Structure preparation

The structures of CB7 and the guest molecules were provided by SAMPL4 organizers, and are shown in Fig. 1. Protonation and tautomer states of the guest molecules at pH 7.4 were determined using Marvin 6.04 software (ChemAxon). In order to address the uncertainty in the protonation of guest **10**, we tested two different protonation states, with the central nitrogen atom carrying a formal charge of +1 or 0. The formal charges of the terminal nitrogen atoms were set to +1 in both cases. The protonation state with +2 net charge, i.e. with the central nitrogen atom deprotonated, exhibited a lower (more favorable) binding free energy using both classical and quantum M2 methods. Therefore, our predicted binding affinities for this guest are based on the +2 charge state.

### Mining minima

The M2 method is described in detail elsewhere [20–22], and so is only briefly summarized here. The host–guest binding free energy is determined from the standard chemical potential ( $\mu^o$ ) of the host, the guest, and their complex, as

$$\Delta G^o = \mu_{\text{complex}}^o - \mu_{\text{host}}^o - \mu_{\text{guest}}^o$$

Based on the predominant states approximation [21], the standard chemical potential of a molecule in solution is approximated as the sum over a collection of  $M$  local energy minima (or energy wells),  $i = 1, \dots, M$ , according to the equation.

$$\mu^o \approx RT \ln \left( \frac{8\pi^2}{C^o} \right) - RT \ln \sum_i Z_i$$

$$Z_i = \int e^{-E(X)/RT} dX$$

where  $R$ ,  $T$ ,  $C^o$ ,  $E(X)$ , and  $Z_i$  are, respectively, the gas constant, the absolute temperature, the standard concentration, the energy as a function of the internal coordinates

$X$ , and the configuration integral over the internal coordinates  $X$  in energy well  $i$ . The energy  $E$  accounts for both the internal energy of the solute and the solvation free energy, both being functions of the internal coordinates. The  $M$  energy minima of a molecule or complex are identified through extensive conformational search. The configurational integral of each local energy minimum is calculated assuming either the quantized rigid-rotor harmonic oscillator (RRHO) approximation [23], for QM/M2, or the classical harmonic approximation with mode scanning (HA/MS) [22] for the classical M2. It is worth remarking, however, that the anharmonicity corrections calculated using mode scanning for host–guest systems are often so small as to be negligible [22]. See Ref. [24] for a detailed description of RRHO approximation, and the associated equations.

### Free energy decomposition

Decomposition of the binding free energy into energetic and entropic components provides additional insight into the driving forces for binding [25]. The ensemble average of an individual energy component is calculated as

$$\langle E \rangle = \sum_i^M P_i E_i; \quad P_i = \frac{Z_i}{\sum_i Z_i}$$

where  $E$  now may be the solute energy, the polar solvation free energy, the nonpolar solvation free energy, or the sum of any of these quantities, for the local energy minimum associated with energy well  $i$ . The weighting factor  $P_i$  is the probability of finding the molecule in the  $i$ th local energy well, which in turn is given by the Boltzmann weight of the  $i$ th energy well. The total configurational entropy of a molecule or complex is calculated as,

$$S^o = \sum_i^M P_i S_i^o - R \sum_i^M P_i \ln(P_i)$$

where  $S_i^o$  is the configurational entropy of the  $i$ th energy well, calculated with the RRHO or HA/MS approximation.

### Classical M2

In the classical M2 method, the potential energy of the molecule is calculated with an empirical force-field energy model, while solvent effects are accounted for using a continuum solvation model. Here, the parameters for bond, angle, torsion and van der Waals parameters were assigned according the CHARMM force field [26], using Discovery Studio Visualizer (Accelrys, Inc.), and atomic partial charges were assigned using the VCharge software [27] (VeraChem, LLC). Initial structures of host–guest

complexes were prepared by docking the guest molecules in the binding site of CB7 using the Autodock Vina program [28].

Binding free energy calculations were performed using the second-generation M2 software, available for download from <http://pharmacy.ucsd.edu/labs/gilson/software1a.html>. The protocol used for performing classical M2 calculations is identical to that reported in Moghaddam et al. [17]. Briefly, each initial structure is subjected to energy minimization using a combination of conjugate gradient and truncated Newton methods, with an energy gradient tolerance of 0.001 kcal/mol/Å. Starting from the energy minimized structure, many local energy minima conformations were identified using the Tork conformational search algorithm [29]. Conformations within 10 kcal/mol of the lowest energy conformation were retained and filtered based on symmetry-corrected root-mean squared distance (RMSD) cutoff of 0.1 Å [30]. Local configuration integrals were computed using the HA/MS correction for the ten softest modes of vibration to account for possible anharmonicity [22]. The Generalized-Born solvation model was used during energy minimization [31, 32], conformational search, and the calculation of configurational integrals, due to its efficiency; and the final solvation free energies are corrected towards Poisson–Boltzmann/surface area (PBSA) solvation model [33, 34], computed for a single representative conformation for each local energy minima. Poisson–Boltzmann calculations were performed by setting the interior and exterior dielectric constants to 1 and 80, respectively, and the dielectric boundary is defined by the Richards molecular surface [35]. Successive iterations of conformational searching were carried out, where each iteration used the most stable conformations identified in the previous iteration as seeds for the conformational search algorithm. This iterative procedure was continued until the free energy converged to within 0.1 kcal/mol.

### Quantum M2

The QM/M2 method is described in detail in a recent publication by our group [19]. In this method, we use the PM6-DH+ semiempirical QM Hamiltonian and the COSMO continuum solvation model as the energy model [36–39]. The initial structures of the host–guest complexes were prepared through manual docking, or by taking the lowest free-energy conformation identified in the classical M2 calculations. We submitted our predictions based on the starting structure that resulted in the lowest (i.e. most favorable) binding free energy. Many low-energy conformations of the host, guest, and complex molecules were generated using the OPLS-2005 force-field and the LMOD conformational search algorithm implemented in the program MacroModel (Schrodinger, LLC.) [40, 41]. A

maximum of 1,000 steps of LMOD conformational search were performed, and conformations within 10 kcal/mol of the lowest energy conformation were retained. Duplicate conformations were filtered based on symmetry and a RMSD cutoff of 0.1 Å. The low-energy conformations generated using the classical OPLS energy model were then refined with the PM6-DH+/COSMO energy model implemented in program MOPAC 2009 using the eigenvector following method until the normalized gradient fell below 0.01 kcal/mol/Å [42]. The COSMO solvation parameters were set to that of water, i.e. a dielectric constant of 78.4 and solvent probe radius of 1.4 Å. The quantum optimized structures were further filtered to remove any duplicates. The configurational integral of each conformation at 300 K, based on the RRHO approximation, was computed using MOPAC's thermochemistry module, without the anharmonicity corrections. The polar solvation free energy determined using COSMO solvation model was supplemented with an additional nonpolar solvation free energy term that is proportional to surface area, to account for cavity formation and van der Waals interactions between the solute and the solvent.

### Empirical corrections to continuum solvation free energies

In a previous study [19], we found that the solvation free energy calculated using a continuum solvation model showed correlation to the errors in calculated binding free energies for CB7 host–guest systems. To correct for this systematic error, we developed CB7-specific empirical corrections for the COSMO and PBSA solvation models; the corrected solvation free energy was computed as

$$\langle W_{corr} \rangle \equiv \alpha \langle W \rangle + \gamma \Delta \langle SASA \rangle + \delta$$

where  $\langle W \rangle$  is the polar solvation free energy calculated using COSMO or PB,  $\Delta \langle SASA \rangle$  is the change in the average surface area of host and guest upon binding, and  $\alpha$ ,  $\gamma$  and  $\delta$  are fitted parameters. The standard values of  $\alpha$ ,  $\gamma$  and  $\delta$ , i.e. without these CB7-specific corrections, are 1.0 (no scaling), 0.006 kcal/mol/Å<sup>2</sup>, and 0.0 kcal/mol (no shift). The value of 0.006 kcal/mol/Å<sup>2</sup> for  $\gamma$  was derived from fitting the experimental hydration free energies of linear alkanes [43]. In addition to using these standard parameter values, we also tried a set of parameters previously optimized specifically for CB7, through fitting of calculated binding affinities to experimental affinities of previously published host–guest complexes [19]. The fitted values of  $(\alpha, \gamma, \delta)$  are (0.9628, 0.0086 kcal/mol/Å<sup>2</sup>, −5.83 kcal/mol), and (1.005, 0.0055 kcal/mol/Å<sup>2</sup>, 3.37 kcal/mol), for the COSMO and PBSA solvation models, respectively. Note that these parameters were determined prior to the SAMPL4 challenge, so the fitting did not include the

present SAMPL4 experimental data. Also, these empirical parameters are likely to be dependent on the associated solute energy model, PM6-DH+ or CHARMM/VCharge, and may not be transferable to other solute energy models.

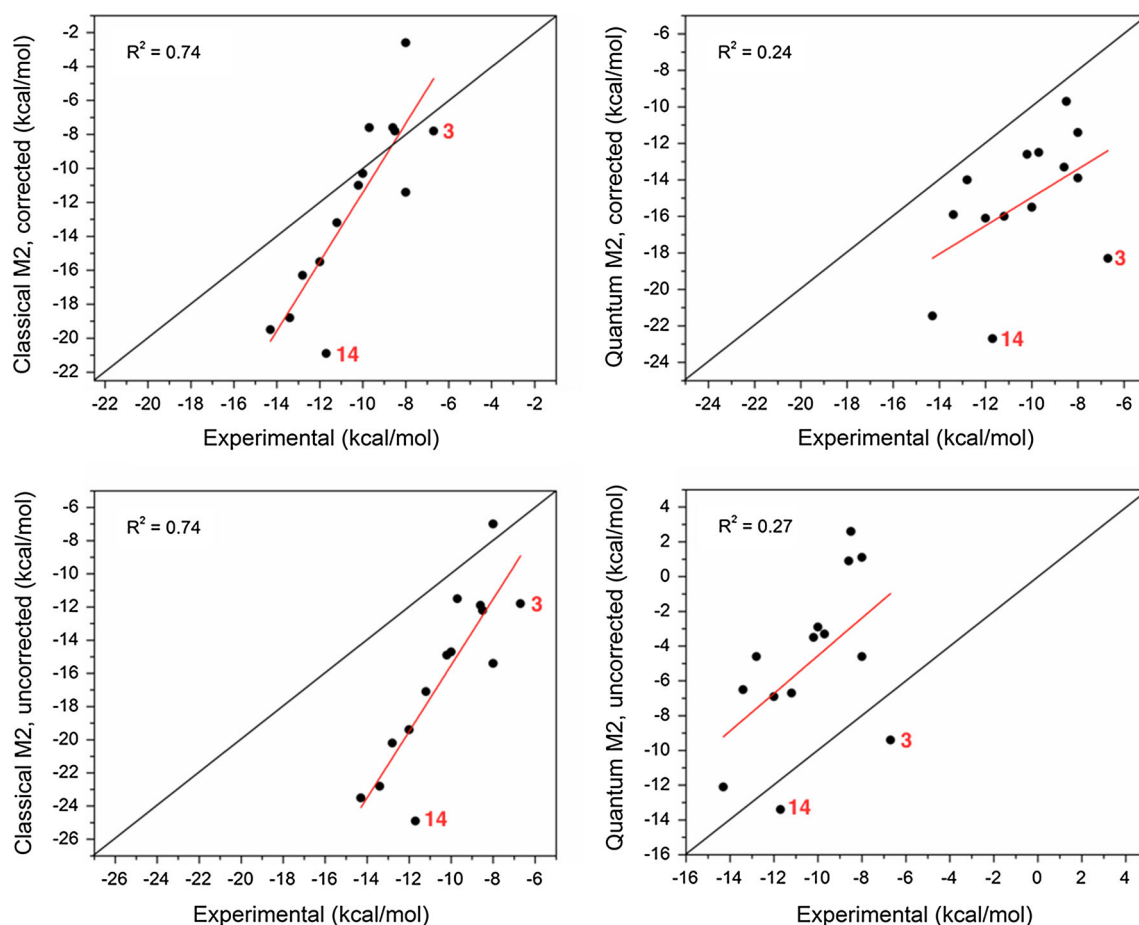
## Results

### Binding free energy predictions

We computed the binding free energies of CB7-guest complexes using classical and quantum M2 methods, both with and without CB7-specific empirical corrections to the continuum solvation free energy. Figure 2 shows a comparison of the calculated binding affinities to the experimental measurements. Binding free energies computed using the classical M2 method correlate well with the experimental free energies, with a correlation coefficient ( $R^2$ ) of 0.74, and a root mean square error (RMSE) of 6.8 kcal/mol. Despite the good correlation between computed and experimental binding affinities, the slope of the

linear regression fit is 2.0, a considerable deviation from the ideal value of 1.0. A similarly large slope was observed in a previous study of CB7 systems with the classical M2 method [16], but not with all force field parameters, and also not for other host-guest systems [12]. Thus, the large slope observed here is not a uniform attribute of the classical M2 method itself. The high RMSE results in part from a constant shift of the binding free energies, as reflected by the mean signed error of  $-5.9$  kcal/mol; i.e. the binding free energies are consistently predicted to be too favorable compared to experiments. Applying the CB7-specific empirical correction to the PBSA solvation model improved the RMSE from 6.8 to 3.9 kcal/mol, but the correlation remained the same at  $R^2 = 0.74$ . To assess the error in relative binding affinities, we calculated RMSE after subtracting the mean unsigned error, to yield the RMSE<sub>o</sub> statistic. This measure of error is 3.5 kcal/mol for the classical M2 method, only slightly lower than the RMSE of the absolute binding affinities.

Although the linear regression slope between the quantum method and experiment is better than that obtained



**Fig. 2** Comparison of M2 predictions with experimental binding free energies. *Left panes* show the results of classical M2 with (*top*) and without (*bottom*) the empirical correction of solvation free energy. *Right panes* show analogous results for quantum M2



using the classical M2 method, 0.7 versus 2.0, the quantum calculations yield a poorer correlation to experiment, with a correlation coefficient of 0.27 and RMSE of 6.7 kcal/mol. Applying the CB7-specific empirical corrections to the COSMO solvation model resulted in only slight improvement of the RMSE from 6.7 to 5.8 kcal/mol, and no improvement in the correlation. The quantum M2 binding free energies without solvation corrections were consistently less favorable than experiment, by 5.4 kcal/mol on average; with the solvation correction, the binding free energies are consistently more favorable than experiment, by 4.9 kcal/mol on average. The poor experimental correlation of the quantum M2 results is largely attributable to two outlying guest molecules, **3** and **14**, whose computed binding free energies deviate from experiment by more than 10 kcal/mol (with solvation corrections), a large error relative to the other guest molecules. Excluding these two outliers, the correlation coefficient was 0.61 and 0.68 and the RMS error was 4.2 and 7.1 kcal/mol, with and without the solvation corrections, respectively. The lack of improvement in correlation with empirical solvation corrections is not totally surprising, as we found in a previous study that these corrections only slightly improve the correlation, but significantly improve the absolute binding free energies [19]. The RMSE<sub>o</sub> of the quantum M2 predictions is 3.2 kcal/mol, somewhat better than that of the classical M2 predictions, owing to the better slope of quantum M2 method.

#### Analysis of guests 3 and 14

No guest molecules emerged as major outliers in the classical M2 calculations, but guests **3** and **14** showed large deviations in the quantum M2 calculations, with deviations in binding free energy of 11.6 and 11.0 kcal/mol, respectively (Table 1), after applying the CB7-specific empirical solvation corrections. Guest **3** is, essentially, *n*-hexane with one terminal hydroxyl and one terminal primary amine group, and guest **14** is a disubstituted adamantane compound. A larger error for guest **14** was somewhat expected, as we previously observed significant deviation between calculation and experiment for a disubstituted adamantane compound with two primary amine groups [19]. As is the case here, the binding free energy was calculated to be significantly more favorable than experiment. Our earlier study also included several mono-substituted adamantane guests, and the calculated binding affinities for these guests show only small deviations from the experimental values. Interestingly, the error for guest **14** in the classical M2 calculations was also relatively high compared to the other guests (Table 1).

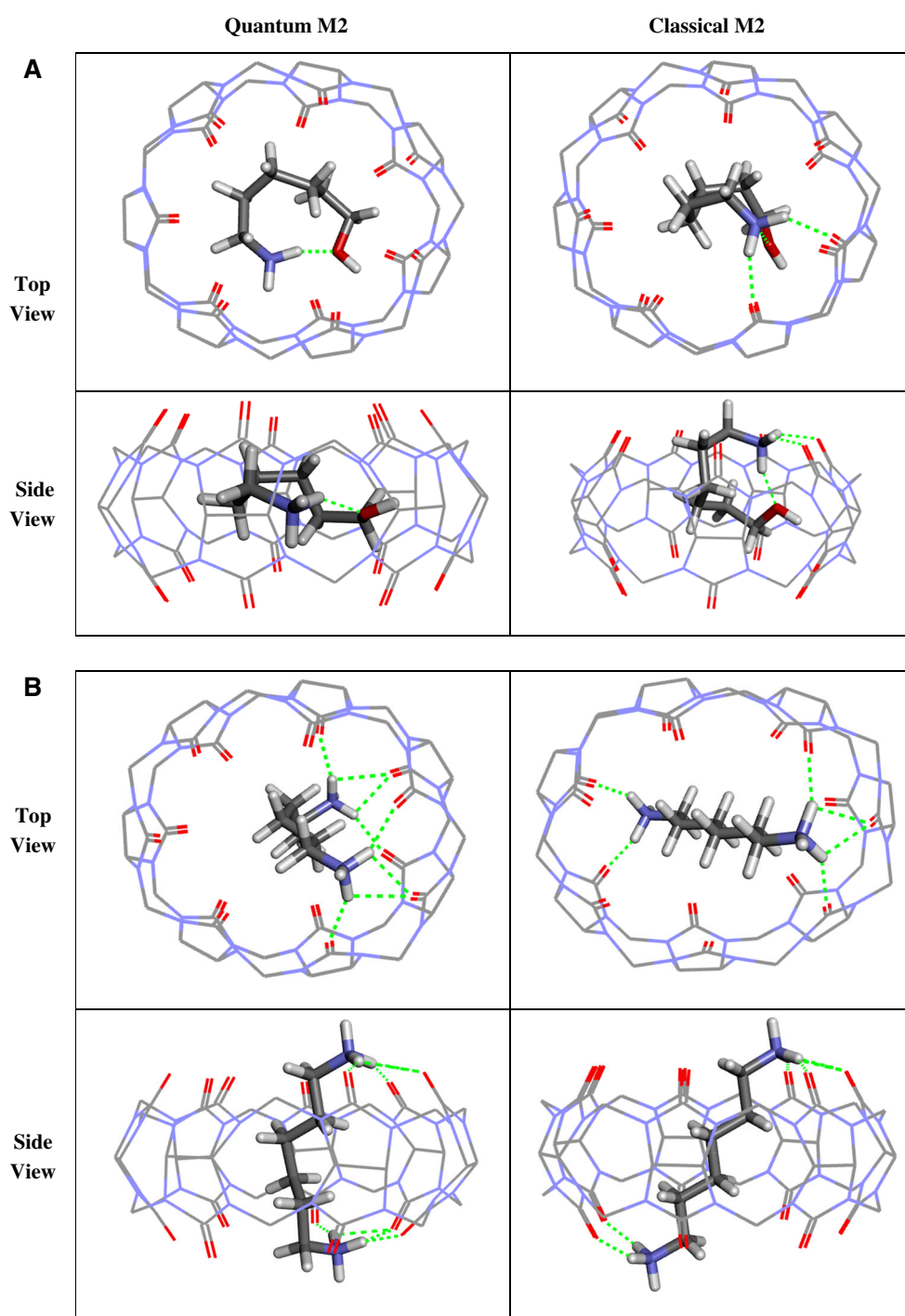
Our prediction for guest **3** using the classical M2 method was significantly better than that using the quantum M2 method. Therefore, we compared the binding modes of guest **3** in the most stable conformations of the complexes generated by the classical and quantum M2 methods (Fig. 3a). In the most stable quantum M2 conformation, the

**Table 1** Summary of computed binding free energy components, including solute energy, solvation free energy, and entropy, from the quantum M2 and classical M2 methods

Guest	$\Delta G_{\text{exp}}$	PM6-DH+/COSMO					CHARMM/Vcharge/PBSA				
		$\Delta G_{\text{calc}}^{\circ}$	$\Delta \langle U \rangle$	$-TS^{\circ}$	$\Delta \langle W_p \rangle$	$\Delta \langle W_{np} \rangle$	$\Delta G_{\text{calc}}^{\circ}$	$\Delta \langle U \rangle$	$-TS^{\circ}$	$\Delta \langle W_p \rangle$	$\Delta \langle W_{np} \rangle$
1	−10.0	−15.5	−187.9	8.9	171.8	−2.5	−10.3	−185.6	13.6	160.8	−2.4
2	−9.7	−12.5	−99.9	6.6	88.3	−1.7	−7.6	−96.1	9.8	77.3	−2.0
3	−6.7	−18.3	−98.5	10.5	77.9	−2.3	−7.8	−105.3	15.3	81.2	−2.3
4	−8.5	−9.7	−178.7	12.5	164.8	−2.5	−7.8	−184.2	16.6	158.9	−2.5
5	−8.6	−13.3	−189.9	11.6	172.5	−1.6	−7.6	−168.2	13.1	146.1	−1.9
6	−8.0	−13.9	−101.8	6.9	88.9	−2.2	−11.4	−100.0	10.4	76.9	−2.1
7	−10.2	−12.6	−99.3	10.0	84.6	−2.1	−11.0	−98.4	10.9	75.1	−2.0
8	−12.0	−16.1	−102.4	9.5	84.8	−2.2	−15.5	−102.7	12.9	73.0	−2.1
9	−12.8	−14.0	−103.4	8.2	89.3	−2.4	−16.3	−107.1	12.9	76.9	−2.3
10	−8.0	−11.4	−184.3	11.6	169.6	−2.5	−2.6	−177.8	11.5	162.6	−2.3
11 <sup>a</sup>	−11.2	−16.0	−102.6	7.0	87.5	−2.1	−13.2	−101.0	10.1	76.4	−2.1
12	−13.4	−15.9	−110.2	12.5	90.1	−2.5	−18.8	−108.8	13.9	75.1	−2.4
13	−14.3	−21.4	−107.6	5.6	88.7	−2.3	−19.5	−107.1	11.2	75.3	−2.2
14	−11.7	−22.7	−110.3	9.5	86.4	−2.5	−20.9	−110.9	13.6	75.3	−2.4

The polar solvation energy term  $\Delta \langle W_p \rangle$  and the non-polar solvation energy term  $\Delta \langle W_{np} \rangle$  shown for both M2 methods are empirically corrected. See “Methods” section

<sup>a</sup> The calculated binding free energy is that of the racemic mixture



**Fig. 3** The most stable conformations of CB7-guest **3** complexes (a) and CB7-guest **4** complexes (b) generated by the quantum M2 and the classical M2 method, shown in two perspectives. Hydrogen bonds are shown as *green dashed lines*

guest molecule adopts a parallel orientation within the host's binding site, with the skeleton of the hexane curled up in the horizontal plane of symmetry of CB7 and both the terminal hydrogen donating groups form hydrogen bonds with the carbonyl groups of the CB7 (Fig. 3a, left). On the other hand, in the most stable classical M2 conformation,

guest **3** adopts a more perpendicular orientation, with the terminal hydroxyl and amine groups forming hydrogen bonds to opposite carbonyl portals of the host (Fig. 3a, right). On the QM energy landscape, this most stable classical M2 conformation is less stable than the most stable quantum M2 conformation by 4.1 kcal/mol. We also

examined the binding modes of guest **4**, an *n*-hexane derivative with two terminal primary amine groups, which is similar to guest **3** in the sense that it also has two terminal hydrogen bond donors and a hydrophobic center. Both classical and quantum M2 predictions for guest **4** were significantly better compared to guest **3**. Binding modes of guest **4** observed in the most stable conformation from classical and quantum M2 methods are somewhat similar (Fig. 3b); interestingly, this guest molecule adopts an extended conformation in both the classical and quantum M2 calculations. Despite the similarity in structure, the free energy of the most stable classical M2 conformation is less stable than the most stable quantum M2 conformation by 5.6 kcal/mol, when it is reevaluated on the QM energy landscape. These results suggest that the large error in computed binding affinity of guest **3** resulted from the inaccuracy of the energy model, rather than from an inadequate conformational search.

#### Comparison of classical and quantum M2 free energy components

For a better understanding of the varying performances of the classical and quantum M2 method, both based on the same idea of M2 and differing only in the energy model, we decomposed the binding affinities of SAMPL4 host–guest complexes computed by both methods into contributions from the solute energy  $\Delta\langle U \rangle$ , the entropy term  $-TS^\circ$ , the polar solvation term,  $\Delta\langle W_p \rangle$  and the non-polar solvation term,  $\Delta\langle W_{np} \rangle$ . The binding free energies and the respective decomposition are summarized in Table 1. Overall the quantum and classical M2 method show similar ranges of values for each individual term, and the favorable contribution to the binding free energy is found always to result from the solute energy and a small contribution from the non-polar solvation energy. The polar solvation free energy and the entropy consistently oppose binding.

The binding free energies computed using classical and quantum M2 methods correlate with each other to some extent,  $R^2 = 0.53$  and  $0.58$ , with and without CB7-specific empirical solvation free energy corrections, respectively (Fig. 4a). We further compared the individual energy components to directly assess the differences in the classical and quantum M2 calculations. For clarity, we present our results based on the calculations with the CB7-specific empirical corrections for solvation free energies, but similar results were obtained even without the solvation free energy corrections. Binding contribution of solute energies, polar solvation energies and nonpolar solvation energies calculated by quantum and classical M2 methods correlate strongly with each other ( $R^2 = 0.97$ ,  $0.97$  and  $0.82$ , respectively), however, this strong correlation is a result of having two

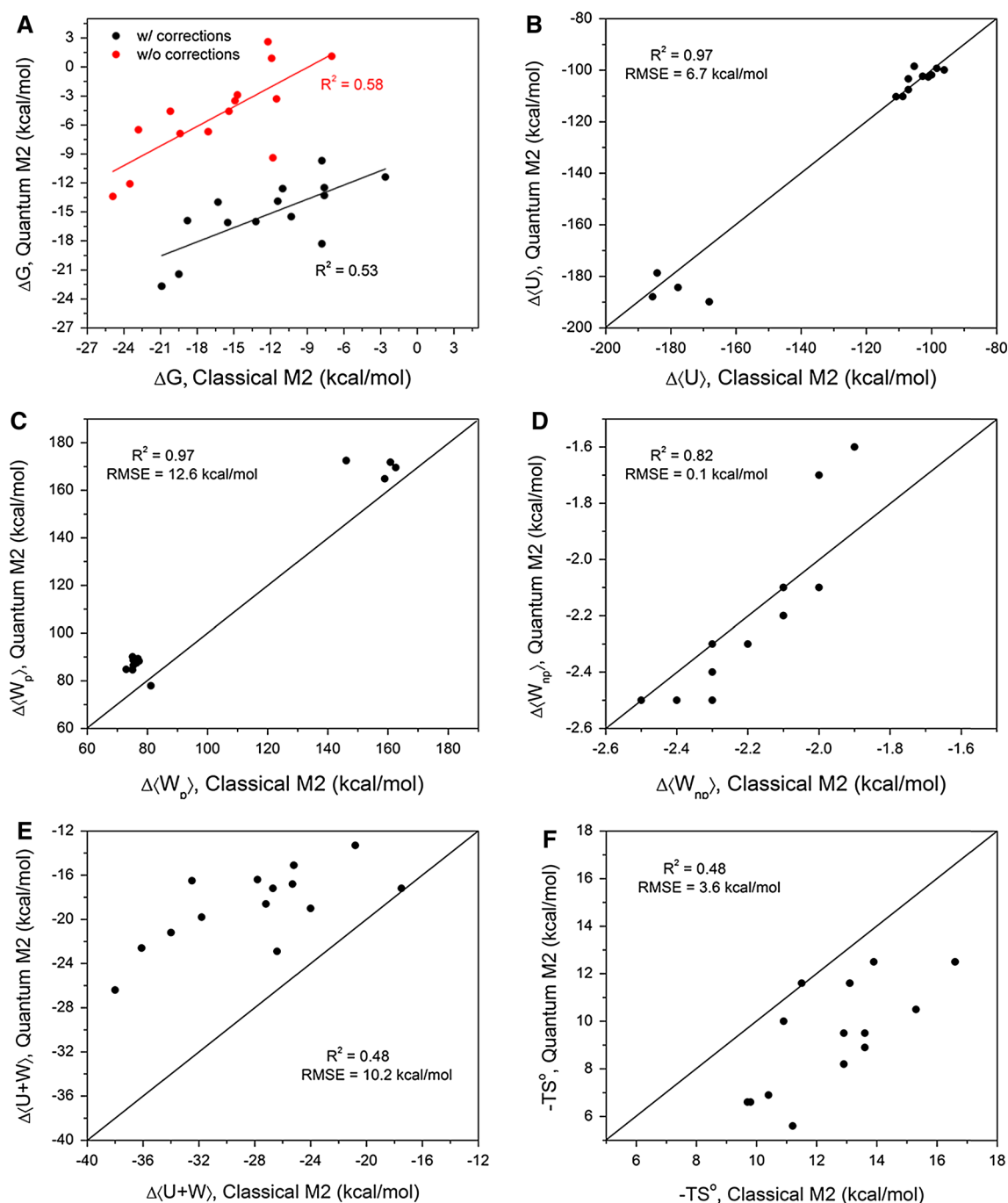
well-separated groups of data points (Fig. 4b–d). The total energy change, i.e. the solute energy change plus the solvation free energy change, showed some correlation between the two methods ( $R^2 = 0.48$ ; Fig. 4e), with an RMS deviation of 10.2 kcal/mol. The RMS deviations of the solute energy change and the polar solvation free energy change, were 6.7 and 12.6 kcal/mol, respectively, suggesting significant differences between the solute Hamiltonians and the solvation models used in the classical and quantum M2 methods, although it is important to keep in mind that the conformations from the two methods are not identical, so the energy comparison is not rigorous. The RMS deviation in the nonpolar solvation free energy was relatively small, 0.1 kcal/mol, which is expected since both the methods use a simple surface area model. The entropic contributions to binding from the classical and quantum M2 calculations were somewhat correlated ( $R^2 = 0.48$ ), with an RMS deviation of 3.6 kcal/mol (Fig. 4f). Overall, these results are consistent with our previous study comparing classical and quantum M2 method using ten CB7 host–guest complexes [19].

#### Driving forces for binding

The decomposition of binding free energy into energetic and entropic components can provide insights into the underlying driving forces for binding [12, 25]. Given the much better performance of the classical M2 method in predicting the binding affinities, we analyze the decomposition of the classical M2 binding free energies (with CB7-specific empirical corrections to solvation free energy) to assess the driving forces of CB7 binding. (Note that the solvation free energy implicitly includes the entropy change of the solvent upon guest binding. As a consequence, the decomposition provided here cannot be directly compared to experiments, for example, isothermal titration calorimetry measurements.) The average change in binding energy (solute energy plus solvation free energy) ranged from  $-17.5$  to  $-38.0$  kcal/mol. Although, solute energy and solvation free energy individually are considerably more negative and positive respectively, the compensation between these two quantities results in relative small and favorable energy changes. The nonpolar solvation free energy contributes a small and nearly constant value of  $-2.2 \pm 0.2$  kcal/mol, towards the binding free energy of the current host–guest systems. The loss in configurational entropy is unfavorable and adds significant free energy cost to the guest binding; the entropy contributions for the present host–guest complexes ranged from 9.8 to 16.6 kcal/mol.

Interestingly, none of the individual binding free energy components correlated well with the experimental binding affinities ( $R^2 < 0.23$ ). On the other hand, the total energy change, i.e. the sum of changes in solute energy and solvation





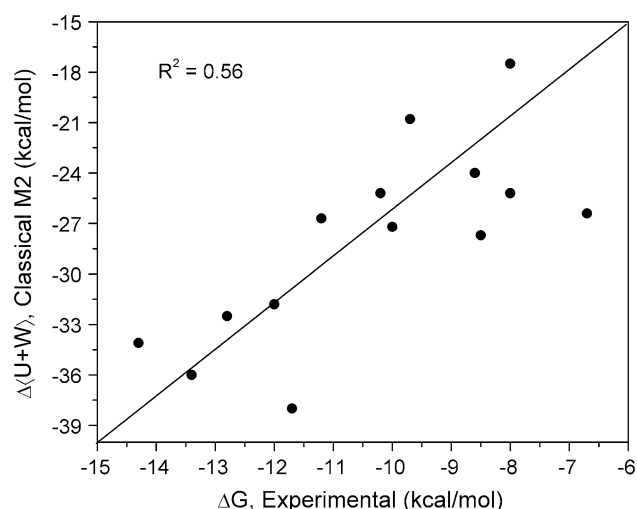
**Fig. 4** Comparison of free energy components, computed by classical and quantum M2 methods

free energy, showed decent correlation ( $R^2 = 0.56$ ) to the experimental binding affinities (Fig. 5).

## Discussion

We used classical and quantum M2 methods to compute the binding affinities of 14 CB7 host–guest complexes in

the SAMPL4 challenge. While the classical M2 binding affinities correlated strongly with experiment, the quantum M2 binding affinities were poorly correlated and suffered from two major outliers. It is unlikely that the errors in the quantum M2 calculations resulted from not finding the most stable conformation in their associated energy landscapes, given that we used aggressive conformational algorithms to enumerate the local energy minima.



**Fig. 5** Comparison of change in total energy (solute energy plus solvation free energy) to experimental binding free energy

Moreover, finding even more stable conformations for these complexes would only shift the calculated binding free energies further from the experimental value.

One of the key approximations made by the M2 method, both classical and quantum, is their use of continuum solvation models. It is not clear how well the conformational distributions of the host, guest, and complex obtained using a continuum solvation model match what would be obtained from an explicit treatment of water molecules [44]. Indeed, it is known that the conformational distributions of proteins are markedly different between explicit-water and implicit water simulations [45, 46]. Nevertheless, our classical M2 calculations showed good correlation with experiment. It is possible that the differences between the classical and quantum M2 methods lie within the different continuum solvation models used, PB and COSMO. Alternatively, the errors in quantum M2 might have resulted from the approximations in the semi-empirical QM Hamiltonian, PM6-DH+. However, it would seem surprising for this energy model to underperform the relatively simplistic force field used in the classical M2 calculations. It is worth noting that another SAMPL4 participant, using a free energy method similar in structure to the quantum M2 method, obtained more accurate results for the CB7 systems in the SAMPL4 host–guest challenge [15]. They used a QM Hamiltonian based on density functional theory (DFT) with a three-body dispersion correction, a continuum solvation model, and a modified RRHO approximation. Further studies are underway to compare PM6-DH+ to DFT methods for computing host–guest binding free energies.

Our predictions for the previous SAMPL3 challenge using the classical M2 method were less than satisfactory

[18]. While CB7 was also included in the SAMPL3 challenge, there were only two guests for this host, so the overall performance of our calculations was dominated by the results for a different host molecule, which posed extra challenges due to its uncertain protonation states and more flexible structure [8]. To avoid the protonation and flexibility issues faced in the SAMPL3 challenge, the experimentalists chose the CB7 host as one of the hosts for SAMPL4. The performance of classical M2 method in the current challenge is more in line with our expectations based on previous studies of CB7 binding, both retrospective and prospective [16, 17].

The binding free energies predicted with both the classical and quantum M2 methods with CB7-specific empirical solvation free energy corrections were consistently more favorable than the measured binding free energies. We speculate that this systematic bias in the predictions might result from an incomplete treatment of the salt effects on binding affinity. Cations in solution bind to the carbonyl portals of cucurbiturils and tend to compete with the binding of guest molecules, thus lowering the affinities of the guests for CB7 [47]. As a consequence, a higher salt concentration will tend to lower these host–guest binding affinities. Our CB7-specific empirical corrections to the solvation free energy were based on previous binding studies of CB7 host–guest complexes measured in a 50 mM sodium acetate (deuterated;  $\text{CD}_3\text{COONa}$ ) buffer or deionized water, whereas the present experimental measurements were obtained in 100 mM sodium phosphate ( $\text{Na}_3\text{PO}_4$ ) buffer [15, 19]. Thus, the empirical fit may have underestimated the affinity lowering effects of the higher salt concentrations in the SAMPL4 experiments. Indeed, CB7 binding affinities of two of the present guests (1 and 4) were previously measured in a 50 mM sodium acetate buffer [48], and the low-salt binding affinities differ from the current measurements by 2.8 and 2.5 kcal/mol, respectively. Thus, the high salt concentration in the present experimental measurements led to higher (less favorable) binding free energies for these guests. This is somewhat sobering, since empirical scoring functions are often parameterized using data from several different sources, which may differ in solvent conditions in a way that could potentially influence the measured binding affinity.

The M2 method provides a detailed breakdown of the binding free energy into solute energy, polar solvation, nonpolar solvation free energy, and configurational entropy. A comparison of each of these components to the experimental binding affinities indicated that none of these free energy components is an independent predictor of binding affinity. However, the solute energy change plus the solvation free energy change showed some correlation to the experimental binding affinity, and so one should

minimally account for these factors in developing approximate computational models of binding. Also, the change in the solute energy plus solvation free energy was found to correlate reasonably well with experiment, substantiating the use of methods such as docking, which are based largely on interaction models which focus on these contributions, to enrich the yield of ligands by initial coarse screens of compound libraries. Nevertheless, accounting for the loss of configurational entropy helps improve the correlation between calculation and experiment.

**Acknowledgments** This study was made possible in part by grant GM61300 from the NIGMS. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Principal Investigator M.K.G. is a founder of and has an equity interest in VeraChem LLC. Although grant GM61300 has been identified for conflict of interest management based on the overall scope of the project and its potential benefit to VeraChem LLC, the research findings included in this particular publication may not necessarily relate to the interests of VeraChem LLC. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies.

## References

- Jorgensen WL (2004) The many roles of computation in drug discovery. *Science* 303(5665):1813–1818
- Gilson MK, Zhou HX (2007) Calculation of protein–ligand binding affinities. *Annu Rev Biophys Biomol* 36:21–42
- Hawkins PCD, Skillman AG, Nicholls A (2007) Comparison of shape-matching and docking as virtual screening tools. *J Med Chem* 50(1):74–82
- Jorgensen WL (2009) Efficient drug lead discovery and optimization. *Acc Chem Res* 42(6):724–733
- Guthrie JP (2009) A blind challenge for computational solvation free energies: introduction and overview. *J Phys Chem B* 113(14):4501–4507
- Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) The SAMPL2 blind prediction challenge: introduction and overview. *J Comput Aided Mol Des* 24(4):259–279
- Geballe MT, Guthrie JP (2012) The SAMPL3 blind prediction challenge: transfer energy overview. *J Comput Aided Mol Des* 26(5):489–496
- Muddana HS, Varnado CD, Bielawski CW, Urbach AR, Isaacs L, Geballe MT, Gilson MK (2012) Blind prediction of host–guest binding affinities: a new SAMPL3 challenge. *J Comput Aided Mol Des* 26(5):475–487
- Skillman AG (2012) SAMPL3: blinded prediction of host–guest binding affinities, hydration free energies, and trypsin inhibitors. *J Comput Aided Mol Des* 26(5):473–474
- Mobley DL, Liu S, Lim NM, Deng N, Branson K, Perryman AL, Forli S, Levy RM, Gallicchio E, Olson AS (2014) Blind prediction of HIV integrase binding from the SAMPL4 challenge. *J Comput Aided Mol Des*
- Snow CD, Sorin EJ, Rhee YM, Pande VS (2005) How well can simulation predict protein folding kinetics and thermodynamics? *Annu Rev Biophys Biomol* 34:43–69
- Chen W, Chang CE, Gilson MK (2004) Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. *Biophys J* 87(5):3035–3049
- Macartney DH (2011) Encapsulation of drug molecules by cucurbiturils: effects on their chemical properties in aqueous solution. *Isr J Chem* 51(5–6):600–615
- Koner AL, Nau WM (2007) Cucurbituril encapsulation of fluorescent dyes. *Supramol Chem* 19(1–2):55–66
- Muddana HS, Fenley AT, Mobley DL, Gilson MK (2014) Blind prediction of the host–guest binding affinities from the SAMPL4 challenge. *J Comput Aided Mol Des*
- Moghaddam S, Yang C, Rekharsky M, Ko YH, Kim K, Inoue Y, Gilson MK (2011) New ultrahigh affinity host–guest complexes of cucurbit[7]uril with bicyclo[2.2.2]octane and adamantane guests: thermodynamic analysis and evaluation of M2 affinity calculations. *J Am Chem Soc* 133(10):3570–3581
- Moghaddam S, Inoue Y, Gilson MK (2009) Host–guest complexes with protein–ligand-like affinities: computational analysis and design. *J Am Chem Soc* 131(11):4012–4021
- Muddana HS, Gilson MK (2012) Prediction of SAMPL3 host–guest binding affinities: evaluating the accuracy of generalized force-fields. *J Comput Aided Mol Des* 26(5):517–525
- Muddana HS, Gilson MK (2012) Calculation of host–guest binding affinities using a quantum-mechanical energy model. *J Chem Theory Comput* 8(6):2023–2033
- Chang CE, Gilson MK (2004) Free energy, entropy, and induced fit in host–guest recognition: calculations with the second-generation mining minima algorithm. *J Am Chem Soc* 126(40):13156–13164
- Head MS, Given JA, Gilson MK (1997) ‘‘Mining minima’’: direct computation of conformational free energy. *J Phys Chem A* 101(8):1609–1618
- Chang CE, Potter MJ, Gilson MK (2003) Calculation of molecular configuration integrals. *J Phys Chem B* 107(4):1048–1055
- Hill TL (1986) An introduction to statistical thermodynamics. Dover, New York
- Zhou HX, Gilson MK (2009) Theory of free energy and entropy in noncovalent binding. *Chem Rev* 109(9):4092–4107
- Fenley AT, Muddana HS, Gilson MK (2012) Entropy–enthalpy transduction caused by conformational shifts can obscure the forces driving protein–ligand binding. *Proc Natl Acad Sci USA* 109(49):20006–20011
- Momany FA, Rone R (1992) Validation of the general-purpose quanta(R)3.2/charmm(R) force-field. *J Comput Chem* 13(7):888–900
- Gilson MK, Gilson HSR, Potter MJ (2003) Fast assignment of accurate partial atomic charges: an electronegativity equalization method that accounts for alternate resonance forms. *J Chem Inf Comput Sci* 43(6):1982–1997
- Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31(2):455–461
- Chang CE, Gilson MK (2003) Tork: conformational analysis method for molecules and complexes. *J Comput Chem* 24(16):1987–1998
- Chen W, Huang J, Gilson MK (2004) Identification of symmetries in molecules and complexes. *J Chem Inf Comput Sci* 44(4):1301–1313
- Qiu D, Shenkin PS, Hollinger FP, Still WC (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J Phys Chem A* 101(16):3005–3014
- Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 112(16):6127–6129
- Gilson MK, Honig B (1988) Calculation of the total electrostatic energy of a macromolecular system—solvation energies, binding energies, and conformational-analysis. *Proteins* 4(1):7–18

34. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin A, Antosiewicz J, Gilson MK, Bagheri B, Scott LR, Mccammon JA (1995) Electrostatics and diffusion of molecules in solution—simulations with the University-of-Houston Brownian dynamics program. *Comput Phys Commun* 91(1–3):57–95
35. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 55 (3):379–IN374
36. Rezac J, Fanfrlik J, Salahub D, Hobza P (2009) Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes. *J Chem Theory Comput* 5(7):1749–1760
37. Korth M, Pitonak M, Rezac J, Hobza P (2010) A transferable H-bonding correction for semiempirical quantum-chemical methods. *J Chem Theory Comput* 6(1):344–352
38. Korth M (2010) Third-generation hydrogen-bonding corrections for semiempirical QM methods and force fields. *J Chem Theory Comput* 6(12):3808–3816
39. Klamt A, Schuurmann G (1993) Cosmo—a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc Perk Trans* 2(5):799–805
40. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still WC (1990) Macro-model—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J Comput Chem* 11(4):440–467
41. Kolossvary I, Guida WC (1996) Low mode search. An efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides. *J Am Chem Soc* 118(21):5011–5019
42. Stewart JJP (1990) Mopac—a semiempirical molecular-orbital program. *J Comput Aided Mol Des* 4(1):1–45
43. Friedman RA, Honig B (1995) A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys J* 69(4): 1528–1535
44. Muddana HS, Sapra NV, Fenley AT, Gilson MK (2013) The electrostatic response of water to neutral polar solutes: implications for continuum solvent modeling. *J Chem Phys* 138(22): 224504
45. Zhou RH (2003) Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins* 53(2):148–161
46. Nymeyer H, Garcia AE (2003) Simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized born approximation with explicit solvent. *Proc Natl Acad Sci USA* 100(24):13934–13939
47. Ong W, Kaifer AE (2004) Salt effects on the apparent stability of the cucurbit[7]uril–methyl viologen inclusion complex. *J Org Chem* 69(4):1383–1385
48. Liu SM, Ruspic C, Mukhopadhyay P, Chakrabarti S, Zavalij PY, Isaacs L (2005) The cucurbit[n]uril family: prime components for self-sorting systems. *J Am Chem Soc* 127(45):15959–15967