J Comput Aided Mol Des (2014) 28:305–317
DOI 10.1007/s10822-014-9735-1

The SAMPL4 host–guest blind prediction challenge: an overview

Hari S. Muddana · Andrew T. Fenley · David L. Mobley · Michael K. Gilson

Received: 14 February 2014 / Accepted: 17 February 2014 / Published online: 6 March 2014

Abstract Prospective validation of methods for computing binding affinities can help assess their predictive power and thus set reasonable expectations for their performance in drug design applications. Supramolecular host–guest systems are excellent model systems for testing such affinity prediction methods, because their small size and limited conformational flexibility, relative to proteins, allows higher throughput and better numerical convergence. The SAMPL4 prediction challenge therefore included a series of host–guest systems, based on two hosts, cucurbit[7]uril and octa-acid. Binding affinities in aqueous solution were measured experimentally for a total of 23 guest molecules. Participants submitted 35 sets of computational predictions for these host–guest systems, based on methods ranging from simple docking, to extensive free energy simulations, to quantum mechanical calculations. Over half of the predictions provided better correlations with experiment than two simple null models, but most methods underperformed the null models in terms of root mean squared error and linear regression slope. Interestingly, the overall performance across all SAMPL4 submissions was similar to that for the prior SAMPL3 host–guest challenge, although the experimentalists took steps to simplify the current challenge. While some methods performed fairly consistently across both hosts, no single approach emerged as consistent top performer, and the nonsystematic nature of the various submissions made it impossible to draw definitive conclusions regarding the best choices of energy models or sampling algorithms. Salt effects emerged as an issue in the calculation of absolute binding affinities of cucurbit[7]uril-guest systems, but were not expected to affect the relative affinities significantly. Useful directions for future rounds of the challenge might involve encouraging participants to carry out some calculations that replicate each others' studies, and to systematically explore parameter options.

Keywords SAMPL4 · Host-guest · Cucurbit[7]uril · Octa-acid · Binding · Prediction · Blind challenge

Introduction

Structure-based computational methods for predicting protein–ligand binding affinities are widely applied in the pharmaceutical industry, to virtually screen and enrich large compound libraries, for early-stage lead identification, and in late-stage lead optimization [1]. However, existing methods still cannot reliably rank candidate compounds by affinity [2], and the goal of predicting protein–ligand binding affinities to within typical experimental uncertainties of perhaps $\sim$ 1 kcal/mol remains elusive. As a consequence, there is still a need for improved techniques for computer-aided ligand design [3].

Methods for predicting protein–ligand binding affinities, such as docking and free energy simulations, are typically

Electronic supplementary material The online version of this article (doi:10.1007/s10822-014-9735-1) contains supplementary material, which is available to authorized users.

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

D. L. Mobley (✉)
Departments of Pharmaceutical Sciences and Chemistry, University of California Irvine, 147 Bison Modular, Irvine, CA 92697, USA
e-mail: dmobley@uci.edu

tuned and tested based on existing binding data. This approach is necessary and effective, but also poses a risk of bias, often unconscious, due to the freedom to tune the protocol and choose the operational parameters (e.g. validation sets, and simulation parameters such as force-field, etc.) to achieve better agreement with experiment. When such tuning is done widely, it can lead to over-fitting of the available data, thereby artificially producing lower errors and inappropriate confidence in the method. As a consequence, retrospectively validated computational methods often perform worse than expected when put to actual use [4].

Prospective, blinded predictions are therefore of particular value for evaluating the predictive power of computational methods. The Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) exercise [5–7], as well as other blind prediction efforts including the Community Structure–Activity Resource [4, 8], Critical Assessment of Protein Structure Prediction [9], and the pKa Cooperative [10] have proven the value of blind prediction challenges in evaluating various computational methods widely used for molecular modeling and drug discovery. Although the first SAMPL challenges focused mainly on prediction of solvation free energies and tautomeric states of drug-like small molecules [5, 6, 11, 12], SAMPL's scope has expanded significantly, through the addition of not only protein–ligand but also host–guest binding affinities, starting with SAMPL3 in 2011 [7, 13] and continuing with the current SAMPL4 exercise [14, 15].

Host–guest systems offer a unique and useful validation approach, as they bridge the scales of the other main SAMPL challenges, i.e. solvation free energies and protein–ligand affinities. Supramolecular hosts are dramatically simpler than proteins, owing to their smaller size (a few hundred atoms) and limited conformational flexibility, and this makes it much easier to generate well-converged computational results that properly test the energy model used [16, 17]. Despite their simplicity, currently available host–guest systems exhibit a wide range of binding affinities in aqueous solution, with maximal affinities rivaling those of the tightest-binding protein–ligand systems [18]. In addition, relative to hydration free energies, host–guest systems offer the advantage of testing the accuracy of not only solute-water but also solute–solute interactions [19]; and, like proteins, but, unlike small molecules in solution, hosts have concave binding surfaces which can lead to locally structured water [20, 21]. For these reasons, host–guest systems are now widely accepted as model systems to validate computational binding affinity methods, and to gain insight into the physical chemistry of molecular recognition.

The current SAMPL4 challenge includes two different host molecules, each with its own set of guests (Fig. 1): cucurbit[7]uril (CB7) [22], a rigid, non-ionizable host; and octa-acid (OA) [23], a somewhat more flexible host, having a hydrophobic pocket and eight ionizable, carboxylic acid groups situated around the entryway and projecting into solvent. Here, we present an overview of the results from all participants in the SAMPL4 host–guest challenge, and discuss issues faced by current computational methods.

Methods

Host–guest systems

The previous SAMPL3 host–guest systems proved more challenging than anticipated, due in large part to uncertainties regarding the protonation states of one host and several of the guest molecules, as well as the possibility of pKa differences in the bound and unbound forms [13]. Accordingly, the participants' results varied greatly depending on the choice of protonation states. The present SAMPL4 systems, especially those involving CB7, were chosen to avoid such issues and thereby focus attention on binding calculations, rather than the related problem of pKa prediction. Figure 1 shows the structures of both host molecules, CB7 and OA. The CB7 host has a relatively nonpolar cavity and two carbonyl-rich portals, and is known to strongly bind guests with a hydrophobic core and one or two cationic groups, which reside near the carbonyls. The unusually high binding affinities, up to −20 kcal/mol, of CB7 for some guests are attributed to the possibility of achieving a high degree of chemical complementarity at a low cost in configurational entropy [18], and to the expulsion of thermodynamically disfavored water from the cavity upon binding [20, 21]. The OA host has a 10 Å deep hydrophobic pocket and an outer surface decorated with eight water-solubilizing carboxylate groups. It is known for its ability to bind hydrophobic or amphiphilic guests in aqueous solution [23], presumably due primarily to the hydrophobic effect, through desolvation of the guest within the hydrophobic pocket of the host. In order to minimize uncertainty in the protonation states of the carboxylate groups on OA, its binding measurements were carried out at pH 9.2 (below). Figure 1 also shows the chemical structures of the CB7 guests, **C1**–**C14**, and the OA guests, **O1**–**O9**.

Drs. Lyle Isaacs and Bruce Gibb generously provided measured binding affinities in advance of publication for the CB7 and OA systems, respectively; these data are summarized in Table 1. The binding affinities for CB7 were measured relative to **C1**, *p*-xylenediammonium, using an ^{1}H NMR competition assay. The binding affinity of guest molecule **C11** was measured for a racemic mixture of the two isomers (*endo* and *exo* isomers). Subsequent

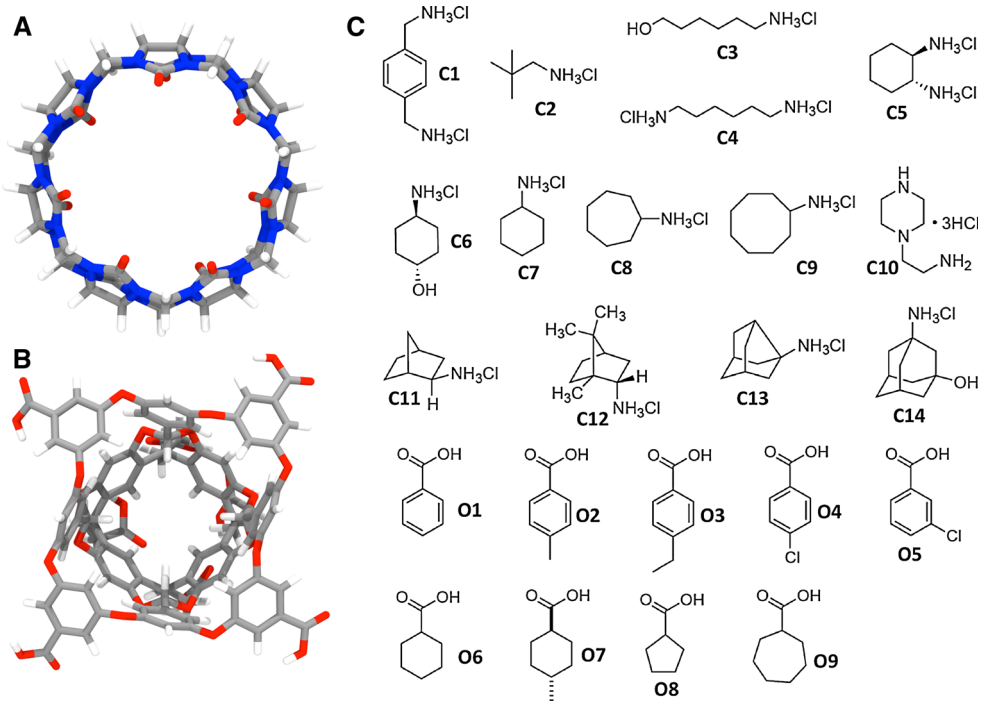


Fig. 1 Structures of cucurbit[7]uril (CB7; **a**) and octa-acid (OA; **b**). **c** Chemical structures of CB7 guest molecules, **C1**–**C14**, and OA guests, **O1**–**O9**. The numbering used here is consistent with that originally provided to the participants

Table 1 Experimental binding affinities of various host–guest complexes

Host	Guest	K_{rel}	K_a (M^{-1})	ΔG (kcal/mol)	Analytic technique
CB7	C1	1.00	$1.7 \pm 0.3 \times 10^{7}$	−9.9 ± 0.1	NMR
CB7	C2	6.23×10^{-1}	$1.1 \pm 0.2 \times 10^{7}$	−9.6 ± 0.1	NMR
CB7	C3	4.20×10^{-3}	$7.1 \pm 1.3 \times 10^{4}$	−6.6 ± 0.1	NMR
CB7	C4	8.77×10^{-2}	$1.5 \pm 0.3 \times 10^{6}$	−8.4 ± 0.1	NMR
CB7	C5	1.06×10^{-1}	$1.8 \pm 0.3 \times 10^{6}$	−8.5 ± 0.1	NMR
CB7	C6	3.82×10^{-2}	$6.5 \pm 1.2 \times 10^{5}$	−7.9 ± 0.1	NMR
CB7	C7	1.40	$2.4 \pm 0.4 \times 10^{7}$	−10.1 ± 0.1	NMR
CB7	C8	2.73×10^{1}	$4.6 \pm 0.8 \times 10^{8}$	−11.8 ± 0.1	NMR
CB7	C9	1.04×10^{2}	$1.8 \pm 0.3 \times 10^{9}$	−12.6 ± 0.1	NMR
CB7	C10	3.43×10^{-2}	$5.8 \pm 1.1 \times 10^{5}$	−7.9 ± 0.1	NMR
CB7	C11	7.38	$1.3 \pm 0.2 \times 10^{8}$	−11.1 ± 0.1	NMR
CB7	C12	3.07×10^{2}	$5.2 \pm 0.9 \times 10^{9}$	−13.3 ± 0.1	NMR
CB7	C13	1.25×10^{3}	$2.1 \pm 0.4 \times 10^{10}$	−14.1 ± 0.1	NMR
CB7	C14	1.81×10^{1}	$3.1 \pm 0.6 \times 10^{8}$	−11.6 ± 0.1	NMR
OA	O1	–	$5.4 \pm 0.4 \times 10^{2}$	−3.73 ± 0.04	NMR
OA	O2	–	$2.0 \pm 0.4 \times 10^{4}$	−5.9 ± 0.1	NMR
OA	O3	–	$4.0 \pm 0.2 \times 10^{4}$	−6.28 ± 0.02	ITC
OA	O4	–	$8.5 \pm 0.4 \times 10^{4}$	−6.72 ± 0.03	ITC
OA	O5	–	$7.1 \pm 0.6 \times 10^{3}$	−5.3 ± 0.1	NMR
OA	O6	–	$1.3 \pm 0.2 \times 10^{4}$	−5.6 ± 0.1	NMR
OA	O7	–	$3.8 \pm 0.4 \times 10^{5}$	−7.6 ± 0.1	ITC
OA	O8	–	$5.4 \pm 0.4 \times 10^{2}$	−3.73 ± 0.04	NMR
OA	O9	–	$7.0 \pm 0.2 \times 10^{4}$	−6.61 ± 0.02	ITC

Measurements are reported as mean ± standard deviation. Uncertainty in binding free energy is calculated through error propagation, $\sigma_{\Delta G} = RT\left(\frac{\sigma_{K_a}}{K_a}\right)$, where R is the gas constant and T = 298 K

NMR nuclear magnetic resonance spectroscopy, *ITC* isothermal titration calorimetry

measurement of the binding affinity of guest **C1** allowed all of the affinities to be reported as absolute, rather than relative, binding free energies. The measurements were performed at 298 K in 100 mM Na_3PO_4 in D_2O, adjusted to pH 7.4 using DCl. Absolute binding affinities of the OA systems were determined using isothermal titration

Table 2 Summary of predictions submitted for the SAMPL4 host–guest binding affinity challenge

ID	Method	Energy model[a]	Solvent model	Conformational sampling	RMSE_r	RMSE_o	R^2	Slope	SAMPL4 ref.
Host: Cucurbit[7]uril (CB7)									
001	NULL1	–	–	–	3.2 ± 0.4	2.2 ± 0.3	0.0 ± 0.0	0.0 ± 0.0	
010	NULL2	–	–	–	3.3 ± 0.4	2.3 ± 0.3	0.2 ± 0.1	0.5 ± 0.2	[45]
187	SIE	GAFF/AM1-BCC	BRI BEM	Wilma	2.7 ± 0.4	1.8 ± 0.3	0.6 ± 0.1	0.18 ± 0.04	[38]
188	SIE + HB	GAFF/AM1-BCC	BRI BEM	Wilma	2.6 ± 0.4	1.8 ± 0.3	0.6 ± 0.1	0.19 ± 0.04	[38]
194[b]	DOCK 3.7	AMSOL	AMSOL	DOCK 3.7	7.9 ± 1.9	5.4 ± 1.3	0.1 ± 0.2	−0.5 ± 0.7	[45]
528	RRHO	DFT-D/HF-3c	COSMO-RS	Manual	3.7 ± 0.7	2.5 ± 0.5	0.8 ± 0.1	1.8 ± 0.3	
541	QM/M2	PM6-DH+	COSMO	Tork	4.5 ± 1.0	3.0 ± 0.7	0.2 ± 0.2	0.7 ± 0.5	[36]
550	M2	CHARMm/VCharge	PBSA	Tork	5.0 ± 1.0	3.4 ± 0.7	0.7 ± 0.1	2.0 ± 0.4	[36]
553	FEP	GAFF/AM1-BCC	TIP4P	Metadynamics	7.2 ± 1.5	4.9 ± 1.0	0.6 ± 0.1	2.2 ± 0.5	
554	FEP	GAFF/AM1-BCC	TIP4P	Metadynamics	5.7 ± 1.2	3.9 ± 0.8	0.6 ± 0.1	1.8 ± 0.4	
555	FEP	GAFF/AM1-BCC	TIP4P-Ew	Metadynamics	8.2 ± 1.4	5.6 ± 1.0	0.3 ± 0.1	1.6 ± 0.6	
556	FEP	GAFF/AM1-BCC	TIP4P-Ew	Metadynamics	6.9 ± 1.2	4.7 ± 0.8	0.3 ± 0.2	1.3 ± 0.6	
557[c]	Enthalpy[d]	GAFF/AM1-BCC	TIP3P/ TIP4P-Ew	MD	4.0 ± 0.6	2.7 ± 0.4	0.7 ± 0.1	1.6 ± 0.3	
558	Enthalpy[d]	GAFF/AM1-BCC	TIP3P	MD	5.8 ± 0.7	4.0 ± 0.5	0.7 ± 0.1	2.2 ± 0.4	
559	Enthalpy[d]	GAFF/AM1-BCC	TIP4P-Ew	MD	4.0 ± 0.7	2.7 ± 0.5	0.6 ± 0.1	1.4 ± 0.3	
560[e]	BAR	AMOEBA	AMOEBA	MD	3.3 ± 0.6	2.2 ± 0.4	0.6 ± 0.1	1.3 ± 0.3	
571	Intuition[f]	–	–	–	6.2 ± 0.7	4.2 ± 0.4	0.3 ± 0.2	1.4 ± 0.5	
574	Water count[g]	GAFF/AM1-BCC	TIP4P	MD	8.6 ± 1.5	5.8 ± 1.0	0.0 ± 0.1	−0.4 ± 0.7	
576	OST	GAFF/AM1-BCC	TIP3P-MOD	MD	2.8 ± 0.4	1.9 ± 0.2	0.8 ± 0.1	1.4 ± 0.2	
577[h]	OST	GAFF/AM1-BCC	TIP3P-MOD	MD	2.8 ± 0.4	1.9 ± 0.3	0.8 ± 0.1	1.4 ± 0.2	
579	PMF	CGenFF	TIP3P	Funnel Metadynamics	5.8 ± 1.1	4.0 ± 0.8	0.1 ± 0.2	−0.4 ± 0.4	[51]
584[e]	BAR	AMOEBA	AMOEBA	MD	3.3 ± 0.6	2.2 ± 0.4	0.6 ± 0.1	1.3 ± 0.3	
600	EES	GAFF/AM1-BCC	TIP3P	MD	5.0 ± 1.1	3.4 ± 0.7	0.7 ± 0.1	1.9 ± 0.4	[52]
601	EES	GAFF/AM1-BCC	TIP3P	MD	4.2 ± 0.7	2.9 ± 0.5	0.6 ± 0.2	1.5 ± 0.3	[52]
Host: Octa-acid (OA)									
002	NULL1	–	–	–	1.9 ± 0.4	1.2 ± 0.3	0.0 ± 0.0	0.0 ± 0.0	
011	NULL2	–	–	–	1.4 ± 0.2	0.9 ± 0.2	0.5 ± 0.2	0.7 ± 0.3	[45]
140	BEDAM	OPLS-AA	AGBNP2	MD	1.3 ± 0.2	0.9 ± 0.1	0.9 ± 0.2	1.5 ± 0.2	
170	FEP	GAFF/AM1-BCC	TIP3P	MD	1.5 ± 0.4	1.0 ± 0.3	0.9 ± 0.1	1.5 ± 0.3	[41]
185	SIE	GAFF/AM1-BCC	BRI BEM	Wilma	1.4 ± 0.3	1.0 ± 0.2	0.7 ± 0.2	0.2 ± 0.1	[38]
186	SIE + HB	GAFF/AM1-BCC	BRI BEM	Wilma	1.4 ± 0.3	0.9 ± 0.2	0.8 ± 0.2	0.3 ± 0.1	[38]
192[b]	DOCK 3.7	AMSOL	AMSOL	DOCK 3.7	1.4 ± 0.3	0.9 ± 0.2	0.7 ± 0.2	0.3 ± 0.1	[45]
193[b]	DOCK 3.7	AMSOL	AMSOL	DOCK 3.7	1.4 ± 0.3	0.9 ± 0.2	0.8 ± 0.2	0.3 ± 0.1	[45]
195[b]	DOCK 3.7	AMSOL	AMSOL	DOCK 3.7	1.4 ± 0.3	0.9 ± 0.2	0.7 ± 0.2	0.3 ± 0.1	[45]
526	FEP	GAFF/RESP	TIP3P	MD	1.3 ± 0.3	0.9 ± 0.2	0.9 ± 0.1	1.5 ± 0.3	[41]
527	RRHO	DFT-D; HF-3c	COSMO-RS	Systematic sampling	5.9 ± 2.2	4.0 ± 1.5	0.4 ± 0.3	2.2 ± 1.6	
551	RRHO	DFT-D; GAFF/RESP	COSMO-RS	MD	8.7 ± 3.9	5.8 ± 2.6	0.5 ± 0.2	3.9 ± 2.0	[41]
552	RRHO	LCCSD(T)/CBS; GAFF/RESP	COSMO-RS	MD	9.2 ± 2.9	6.1 ± 1.9	0.4 ± 0.2	3.3 ± 2.1	[41]
578	NB-BAR	DFT-D; GAFF/RESP	QM/MM	MD	6.5 ± 1.0	4.4 ± 0.7	0.0 ± 0.2	0.4 ± 1.3	[41]

Table 2 continued

ID	Method	Energy model[a]	Solvent model	Conformational sampling	RMSE_r	RMSE_o	R^2	Slope	SAMPL4 ref.
602	EES	GAFF/AM1-BCC	TIP3P	EES	1.9 ± 0.4	1.3 ± 0.3	0.8 ± 0.1	1.7 ± 0.4	[52]

Null1 null model with a constant value assigned to all guests (0.0 and −5.0 for CB7 and OA, respectively), *Null1* null model based on number of heavy atoms (1.5 kcal/mol per heavy atom [46]), *SIE* solvated interaction energy [39]; *SIE + HB* solvated interaction energy model with explicit corrections for hydrogen bonding [38, 40], *DOCK* UCSF docking program [53], *RRHO* rigid rotor harmonic oscillator approximation [35], *QM/M2* quantum mechanical mining minima [37], *M2* mining minima [17], *FEP* free energy perturbation [54, 55], *BAR* Bennett acceptance ratio [32], *OST* orthogonal space tempering [29], *PMF* potential of mean force [56], *EES* expanded ensemble simulations [57], *BEDAM* binding energy distribution analysis method [42], *NB-BAR* non-Boltzmann Bennett acceptance ratio [58], *GAFF* general amber force field [26], *AM1-BCC* [27, 28], *AMSOL* a semiempirical quantum chemistry program [59], *DFT-D* dispersion-corrected density functional theory [60], *HF-3c* corrected small basis set Hartree–Fock method [61], *PM6-DH +* PM6 semiempirical method with dispersion and hydrogen bonding corrections [62, 63], *CHARMm* chemistry at Harvard molecular mechanics [64]; *VCharge* VeraChem charge model [65], *AMOEBA* atomic multipole optimized energetics for biomolecular simulation [33], *CGenFF* CHARMM general force field [66], *OPLS-AA* optimized potentials for liquid simulations all-atom force field [43], *RESP* restrained electrostatic potential [67], *LCCSD(T)* local coupled-cluster method with single and double excitations, and non-iterative perturbative treatment of triple excitations [68], *CBS* complete basis set, *BRI BEM* biotechnology research institute boundary element method for solving Poisson equation [69], *COSMO-RS* conductor-like screening model for real solvents [70]; *COSMO* conductor-like screening model [71], *PBSA* Poisson-Boltzmann surface area [72], *TIP4P* transferable interaction potential four-point [30], *TIP4P-Ew* TIP4P for use with Ewald summation [49], *TIP3P* transferable interaction potential three-point [30], AGBNP2 analytical generalized Born plus non-polar (version 2) [44], *QM/MM* mixed quantum mechanics and molecular mechanics, *Wilma* an exhaustive docking algorithm [38], *Tork* an aggressive conformational search method using normal modes [73], *Metadynamics* [74], *MD* molecular dynamics, *Funnel Metadynamics* [75]

[a] In case of mixed energy models used in the calculations, the different energy models are listed separated by a semicolon

[b] Per the submitters, submissions 192–195 are potentially flawed due to problems with the parameter files not being processed correctly

[c] The free energy of binding was estimated as the average of submissions 558 and 559

[d] Direct calculation of enthalpy change from long MD simulations of end states

[e] Submissions 560 and 584 are identical except for the reference binding affinity assigned for guest **C1**

[f] Submitter provided numerical estimates based on his/her chemical intuition regarding these systems

[g] Binding affinities were estimated based on change in the number waters within 4.8 Å of the guest between the bound and the unbound states

[h] The approach used in submissions 576 and 577 is identical, except that in submission 577, the participant used ionic-strength dependent binding affinities of CB7 published in an earlier work [76] to calibrate AMBER Na+ parameters

calorimetry (ITC) or ^{1}H NMR spectroscopy, for strong and weak binders, respectively. The measurements were performed at 298 K in 10 mM sodium tetraborate ($Na_2B_4O_7$) buffer at pH 9.2. In all cases, a 1:1 stoichiometry of binding was confirmed. Detailed descriptions of the experiments are reported elsewhere [24, 25].

Challenge design and logistics

The SAMPL4 challenge for host–guest binding affinities was designed similarly to the previous SAMPL3 challenge [13]. While past SAMPL challenges were primarily organized by OpenEye Scientific Software, this year the challenge was coordinated by academic researchers, although OpenEye provided significant logistical support. The challenge was advertised by posting it on the SAMPL website (http://sampl.eyesopen.com), as well as through e-mails to past participants, the computational chemistry list (http://www.ccl.net), and a number of researchers in the field. These announcements were sent out beginning in January 2013, and the participants then signed up through the SAMPL website. Registered participants were notified upon release of the host–guest challenge systems in mid-February, and submissions were due August 16. Experimental results were released to participants as soon as possible after the Aug. 16 deadline, and our analysis of the performance of the different submissions was released as completed. The overall outcomes of the challenge were discussed at a dedicated workshop on September 20, 2013, at Stanford University. All participants were invited to submit manuscripts for the present special issue in the Journal of Computer-Aided Molecular Design.

The mid-February release included information on the experimental methods used to measure the binding affinities, including the buffer composition and pH (but see below); crystallographic structures of the hosts, as MOL2 files; and 2D and computationally modeled 3D structures of the guest molecules, in the form of 2D drawings, SMILES strings, and MOL2 and SDF files. Geometries of both the isomeric forms of guest **C11** were provided to the participants, for convenience. Although both the buffer composition and pH used in the CB7 measurements were provided in the initial release in February, only the pH was provided at that time for the OA measurements, and so the

participants were not informed of the buffer composition used in the OA measurements. Participants were furthermore informed that the protonation, conformation, and tautomer states provided might not be optimal for the experimental conditions, and were given suitable disclaimers about conformational sampling and other factors. A document containing background information on the hosts, and references to relevant publications, was also provided. For the CB7 host, our initial understanding was that the results would be reported as binding free energies relative to guest **C1**, *p*-xylenediammonium. Participants were therefore instructed to submit binding free energies relative to this compound **C1**. For the OA system, participants were given a special disclaimer about the uncertain protonation state of the host. For both hosts, participants were asked to submit the uncertainties of their predictions, in addition to the binding free energies themselves.

Error analysis

All submissions were analyzed by standard error metrics, including root mean-squared (RMS) error, Pearson coefficient of determination (R^2), and linear regression slope. As noted above, we had anticipated that the CB7 affinity measurements would yield relative free energies, so participants were asked to predict relative binding free energies. In the end, however, the experiments additionally yielded absolute binding free energies. Because of an unclear file format specification, many participants submitted relative binding free energy predictions, but some submitted absolute predictions. At first, we handled this by converting all CB7 relative free energy predictions into absolute values by adding the experimental binding free energy of the reference compound, **C1**. However, this initial approach proved problematic, as follows. One group had submitted both absolute and relative free energy predictions based on the same set of absolute free energy calculations. Because their calculated absolute free energy for compound **C1** happened to be particularly poor, when they converted these free energies into relative free energies by subtracting their computed result, and we then converted them back into absolute free energies by adding the experimental result for **C1**, this resulted in poor apparent performance. Effectively, our procedure resulted in the difference between their predicted absolute binding free energy for **C1** and the actual experimental value being added as an offset to all the absolute free energies analyzed. Thus, our initial approach of converting relative free energies into absolute free energies by adding the experimentally measured binding affinity of **C1** turned out to penalize any submission that happened to perform particularly poorly for compound **C1**. To avoid this problem, we computed the RMS error of predicted binding affinities after subtracting the average signed error, which is referred to as RMSE_o in the following, given by,

$$RMSE\_o = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\Delta G_i^{\exp} - \Delta G_i^{calc} - \frac{1}{n}\sum_{j=1}^{n}\left(\Delta G_j^{\exp} - \Delta G_j^{calc}\right)\right]^2}$$

where n is the number of measurements, ΔG^{exp} and ΔG^{calc} are the experimental and calculated binding affinities, respectively. For similar reasons, we evaluated the accuracy of relative binding free energies by considering all differences among all pairs of guest molecules; the resulting metric is termed RMSE_r, given by,

$$RMSE\_r = \sqrt{\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left[\left(\Delta G_j^{calc} - \Delta G_i^{calc}\right) - \left(\Delta G_j^{\exp} - \Delta G_i^{\exp}\right)\right]^2}$$

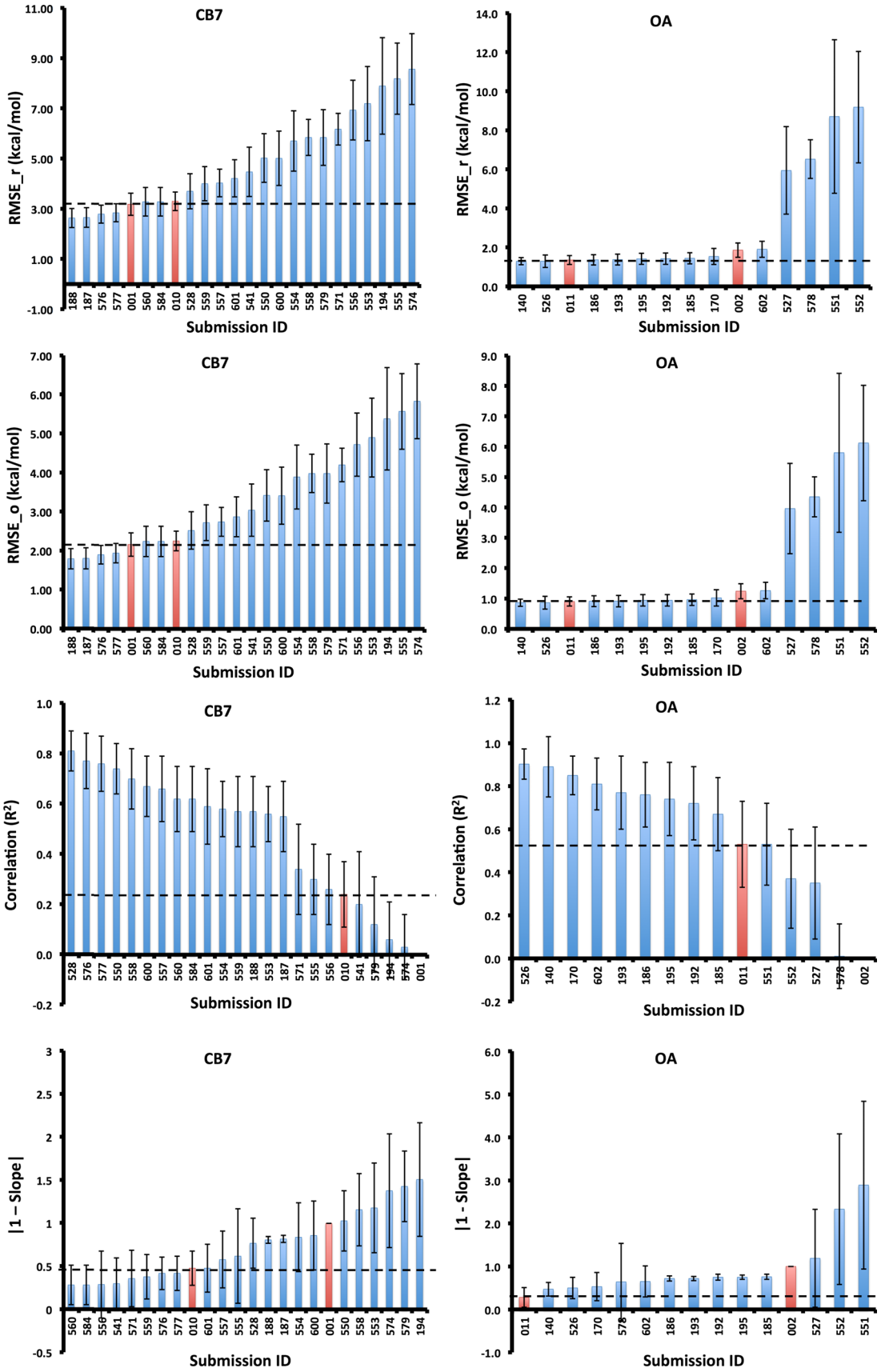
The uncertainties of the various error metrics were determined with a bootstrapping procedure which accounted for experimental uncertainty; details are provided in another article in this issue [14].

Results

The current round of the SAMPL host–guest binding affinity challenge saw an increase in the number of submissions relative to the previous SAMPL3 host–guest challenge, with 22 submissions for the CB7 host, and 13 submissions for the OA host, compared with 18 submissions in toto for SAMPL3 [13]. Table 2 summarizes the various submissions, along with various measures of accuracy. Scatter plots of the predicted versus experimental binding affinities of all the submissions are given in supplementary material. Table 2 also includes results for two simple null models for each of the hosts, as detailed below.

The methods used by the participants ranged from simple scoring methods, to free energy calculations using extensive conformational sampling, to quantum mechanical energy models based on a relatively modest number of conformations. The methods differed from each other in many ways, including the solute energy model (empirical force field or quantum mechanical theory), solvent model (continuum or explicit water), and sampling method

Fig. 2 Rank ordering of submissions based on various error metrics ▶ for the CB7 and OA systems. Null models are shown in *red* and *horizontal dotted lines* are provided as visual guides to distinguish which submissions surpass the null models


CB7
OA
RMSE_r (kcal/mol)
RMSE_o (kcal/mol)
Correlation (R²)
|1 – Slope|
|1 - Slope|
Submission ID

(docking, simulation, and conformational search). Free energy simulations with classical force fields were particularly popular, and, as in SAMPL3, the general AMBER force field (GAFF) [26] coupled with AM1-BCC [27, 28] partial charge model remained the predominant choice for the assignment of force field parameters. A significant number of submissions, 19 of 35, used an explicit-water model to include solvent interactions, a marked departure from SAMPL3 where implicit solvation models were the most popular. The number of submissions based on ab initio quantum mechanical theory also increased relative to the SAMPL3 challenge.

Overview of results

The correlation coefficients (R^2) of the various methods ranged from near zero to greater than 0.8, for both host datasets (Fig. 2). Overall, the RMSE_o and RMSE_r measures for the CB7 host are higher than those for OA host, perhaps due to the broader range of CB7 binding affinities. The majority of submissions, 23 of 35, had a linear regression slope greater than 1.0.

While none of the methods ranked best across all the error metrics for the CB7 dataset, a few alchemical free energy methods, using classical force fields, were consistently among the top for all the metrics. The orthogonal space tempering (OST) method [29] (ID 576, 577) yielded a correlation coefficient of 0.8, along with RMSE_o and RMSE_r values of 1.9 and 2.8 kcal/mol, respectively. This method used GAFF/AM1-BCC with a modified TIP3P water model [30, 31], coupled with an enhanced conformational sampling method. Another method, which used the standard Bennett Acceptance Ratio (BAR) free energy method [32], coupled with a more sophisticated, polarizable force field, AMOEBA [33, 34] (IDs 560, 584), did nearly as well, with a correlation coefficient of 0.6, and RMSE_o and RMSE_r values of 2.2 and 3.3 kcal/mol, respectively. A few end-point methods also did as well as the above alchemical methods. A method using density functional theory with three-body dispersion corrections (DFT-D3) and the COSMO-RS continuum solvation model [35], coupled with a rigid-rotor harmonic oscillator-like approximation for configurational entropy (ID 528) had a correlation coefficient (R^2) of 0.8 with experiment. The RMSE_o and RMSE_r metrics for this method were also fairly low, at 2.5 and 3.7 kcal/mol, respectively. Interestingly, this approach yielded below average accuracy for the OA host (ID 527), with a low correlation coefficient of 0.4. The poor performance of this method for OA host may have resulted from the high flexibility of OA, as the participant used single optimized structures for computing binding affinities. Furthermore, the participant assumed OA to be fully protonated even at pH 9.2, which seems rather unlikely. Another method, mining minima (M2) is structurally similar [17], in that it estimates free energies based on what is essentially a harmonic approximation, for a solute potential energy model combined with an implicit solvent model. This approach, using a classical force field (ID 550), also performed well for CB7, in terms of correlation ($R^2 = 0.7$), but had higher RMSE_o and RMSE_r values of 3.5 and 5.0 kcal/mol, respectively, and a large linear regression slope of 2.0 [36]. Perhaps surprisingly, the M2 method when coupled with a semi-empirical quantum mechanical energy model [36, 37] (ID 541) did not perform as well as the classical version. Interestingly, the relatively simple and fast SIE method [38, 39] (IDs 187, 188), which is based on docking using a classical force field and a continuum solvation model, performed slightly better than average in terms of correlation ($R^2 = 0.6$, both with and without hydrogen bond corrections [40]), and afforded the best RMSE_o and RMSE_r measures of 1.8 and 2.6 kcal/mol, respectively. On the other hand, its linear regression provided particularly shallow slopes (0.18–0.19).

For the OA host, predictions based on free energy simulations with non-polarizable classical force fields again performed well (IDs 170, 526, and 602). The BAR method coupled with a classical force field and explicit water model [41] (ID 170, 526) performed the best in terms of correlation coefficient, RMSE_o, and RMS_r metrics, closely followed by the BEDAM method [42] coupled to the OPLS-AA classical force field [43] and the AGBNP2 [44] continuum solvation model (ID 140). Curiously, all the methods based on quantum mechanical energy models (IDs 527, 551, 552, and 578) performed below average for the OA host. We note that many methods performed well for the OA host, and the differences among them are not statistically significant (See supplementary material). Unfortunately, there was little overlap between the methods tested for CB7 and for OA. For example, the M2, OST and AMOEBA approaches were tested for CB7 but not OA. Given the inconsistent performances of those methods that were tested across both hosts, it is difficult to draw conclusions about which methods are most accurate in general.

Comparison with null models

In order to assess the significance of the present results, we put them into the context of two simple null models [45]. One assigned a constant binding affinity to all the guests, and the other assigned -1.5 kcal/mol of binding free energy per heavy atom [46]. For CB7, both null models had rather low RMSE_o and RMSE_r values of $\sim$2.3 and $\sim$3.3 kcal/mol, respectively, and the heavy atom count model showed some correlation, with $R^2 = 0.2$. For OA host, both the null models had low RMSE_o and RMSE_r,

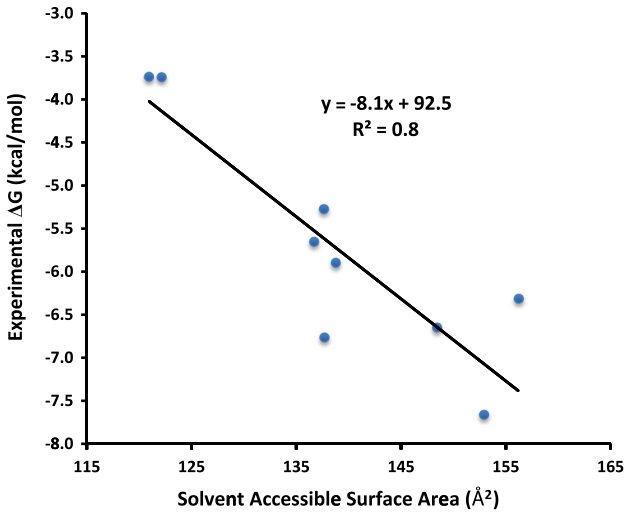


Fig. 3 Binding free energy of guests for the OA system as a function of guest's solvent accessible surface area

between 0.9 and 1.9 kcal/mol, which can be explained by the narrow binding affinity range (3 kcal/mol) for these systems. In terms of correlation, the null model based on heavy atom count did better for the OA host, with an R^2 of 0.5. For the OA host, binding is predominantly due to hydrophobic effect [23], as reflected here by the fact that solvent-accessible surface area of the guests correlates strongly with the binding affinity, with $R^2 = 0.8$ (Fig. 3). This possibly explains why the heavy atom count, which is presumably proportional to the solvent-accessible surface area for small molecules, correlated well with the experiment.

For CB7, 18 of the 22 submissions outperformed both null models in terms of correlation, but fewer than half performed equally well or better in terms of RMSE_o, RMSE_r, and linear regression slope (Fig. 2). The methods that generally performed better than both the null models are: 1) SIE (IDs 187, 188), OST (IDs 576, 577), and BAR/AMOEBA (IDs 560, 584). Interestingly, both SIE and OST performed well in SAMPL3 as well [13]. The third method, BAR with the polarizable force field AMOEBA, is a new entry this year. Perhaps surprisingly, these three methods differ from each other in many ways: as described above, OST is an enhanced sampling free energy perturbation approach using a pairwise additive force field; BAR/AMOEBA is a standard free energy perturbation method using a polarizable force field; and SIE is a fast docking model with a classical force field and continuum solvent model.

On the other hand, for the OA host, many of the methods performed equally well or better than both null models, in terms of both the correlation and the RMS error (Fig. 2). Interestingly, the methods that did significantly worse than the null models are all based on quantum mechanical energy models, including the method (ID 527) that did the best on the CB7 host in terms of correlation (above). It is also worth noting that the SIE method did better than the null models for both CB7 and OA hosts, and also did better than most other methods in the previous SAMPL3 challenge [13]. The BEDAM method with AGBNP2 implicit solvent model, one of the top-performing methods for the OA host, also did well in the SAMPL3 challenge [13]; however, it was not used for the present CB7 systems. It is encouraging to see that some methods are performing well rather consistently, even if no single approach emerges as a consistent top performer.

SAMPL4 versus SAMPL3

It is of interest to know how well the different computational methods performed this round compared to the last. In terms of correlation, the overall performance this year (average $R^2 = 0.56$) is essentially the same as that of the previous SAMPL3 challenge [13] (average $R^2 = 0.54$). The performance for each of the hosts, CB7 and OA, is not significantly different, a departure from SAMPL3, where the performance for the CB7/CB8 hosts was significantly better than the performance for the acyclic cucurbituril congener host (Host1 in SAMPL3), although the number of guests for CB7/CB8 were low in number. The OA host was expected to be somewhat less challenging than Host1 in SAMPL3 [13], since the protonation state of OA is expected to be well defined based on the experimental pH (9.2), and moreover, the ionic carboxylate groups are located away from the binding cavity. Accordingly, the performance of the different methods for OA host was considerably better than that for the SAMPL3 Host1. Based on SAMPL3, we also expected the blind predictions for CB7 in SAMPL4 to perform about as well as they did for the CB7/CB8 hosts in SAMPL3, and this is at least roughly true. Thus, the average R^2 for all the submissions is equal to 0.50, with 15 submissions performing better than the average; while, for CB7 and CB8 in SAMPL3, the average R^2 value was 0.69. Thus, the overall performance of this round is similar to the previous one.

Discussion

The participation of computational groups in the SAMPL4 host–guest binding affinity challenge has increased since the last round, indicating a growing interest in blind evaluation of computational methods based on host–guest data. The methods tested in the present challenge are quite varied, and each focusing on certain aspects of the binding affinity calculation. For example, some focused on using

presumably more accurate ab initio or semiempirical quantum mechanics, whereas others used simpler classical force fields, and focused instead on conformational sampling. Although this study is not sufficiently large and systematic to allow one to draw definite conclusions regarding the usefulness and accuracy of any specific methodological choice (e.g. explicit vs. continuum solvent, or quantum vs. classical), the various submissions offer some interesting insights.

Previous studies have shown that explicit water effects might play an important role in binding for the CB7 system [20, 21, 47, 48], and suggested that expulsion of thermodynamically disfavored structured waters in the host's cavity might drive the high binding affinities of some guests for CB7. Nonetheless, the overall performance has been similar for both continuum and explicit-water solvation models, and some of the better performing methods here used continuum solvation models. Currently, there is no clear indication as to which solvation model is better, at least when applied to these host–guest systems. It is important to note that the overall predicted affinities depend on a number of other important factors besides the solvent model, including the force field and conformational sampling methods. Errors due to these other factors could be larger and therefore mask the importance of a specific solvation model.

A recurring theme in the host–guest binding affinity challenge is that different free energy methods using very similar energy models (force field, partial charges, and water model) sometimes yield significantly different predictions. For example, submission IDs 553-556 used an energy model very similar to the OST submissions (ID 576, 577), but generated significantly different results. Although the small variations across the water models used (TIP3P [30], TIP3P-Mod [31], TIP4P [30], TIP4P-Ew [49]) might cause these large variations in RMS errors and correlation coefficients, two submissions (IDs 558 and 559) that differed only by water model showed similar performance, suggesting that the variations may result from something other than the water models. One may speculate that these differences could stem from subtle methodological differences, such as the treatments of ions, periodic electrostatics, convergence of conformational sampling, or issues in the manual part of calculation setup. On the other hand, the results were very similar for BAR calculations of OA-guest binding with two different charge models, RESP and AM1-BCC (IDs 170 and 526).

Finally, it is worth noting that the effect of dissolved salt on binding was a potential issue for both hosts, since the measurements were done in salt solutions. Amongst methods based on molecular dynamics simulations, a few submissions accounted explicitly for salt concentration effects, while others only included counterions to maintain neutrality of the system. While OA binding affinity measurements were done at a relatively low salt concentration, 10 mM sodium tetraborate, the CB7 experiments were carried out in 100 mM sodium phosphate buffer. Cations in solution bind to the carbonyl portals of cucurbiturils and compete with the guest molecules for binding, thus lowering the absolute binding affinities [50]. Thus, binding affinities of guests **C1** and **C4** previously measured at a relative low salt concentration (50 mM NaO_2CCD_3) [22] were more favorable by 2.8 and 2.4 kcal/mol, respectively than the newly measured SAMPL4 values at higher salt concentration. The difference in relative binding free energy is thus reduced by 0.4 kcal/mol in the current study, relative to the prior results obtained under low salt conditions. Furthermore, all of the CB7 guests here are cationic and therefore could strongly interact with the counterions in solution, further lowering the guest's binding affinity. Salt concentration can also be important for OA binding calculations, since the host is expected to have a net charge of −8e at pH 9.2. However, the OA binding measurements were done at relatively low salt concentration, so salt effects might be less important than in the case of CB7 binding.

Conclusions and perspectives

The participation of computational groups in the host–guest challenge has grown significantly since SAMPL3, and a wide range of methodological approaches is now represented. The present results offer a uniquely objective snapshot of the state of the art in affinity predictions for these simple model systems, and allow some insights to be extracted regarding the strengths and weaknesses of the various approaches. On one hand, some approaches appear to be emerging as fairly reliable contenders across the systems of SAMPL3 and SAMPL4. On the other hand, there are still decided inconsistencies in the pooled results. For example, high accuracy for one host need not imply high accuracy for another; the most accurate results for a given system may derive from strikingly varied approaches; and quite similar approaches may yield surprisingly different results. In the future, it may be useful to try building in basic consistency checks by encouraging multiple submissions with overlapping methods. In addition, more systematic combinations of methodological components (e.g. free energy methods and water models) could enable clearer answers to such questions as which energy models are most accurate and the role of conformational sampling. Given the collaborative spirit of this ongoing exercise, it may indeed be possible to integrate some level of coordination along these lines in the coming years.

Acknowledgments We thank OpenEye software for providing logistical support in the form of web hosting, technical support, and some financial support for the SAMPL4 workshop. We also thank Jay Ponder (Washington University) and Teng Lin (Schrödinger) for helpful discussions about analyzing absolute versus relative free energy predictions, and Vijay Pande for hosting the workshop at Stanford University. DLM acknowledges the financial support of the National Institutes of Health (1R15GM096257-01A1) and appreciates the support of the GreenPlanet computing facility at UC Irvine, supported in part by NSF CHE-0840513. MKG acknowledges funding from National Institute of General Medical Sciences (GM61300). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS, NIH, or NSF.

References

1. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3(11):935–949
2. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49(20):5912–5931
3. Gilson MK, Zhou H-X (2007) Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct 36:21–42
4. Damm-Ganamet KL, Smith RD, Dunbar JB Jr, Stuckey JA, Carlson HA (2013) CSAR benchmark exercise 2011–2012: evaluation of results from docking and relative ranking of blinded congeneric series. J Chem Inf Model 53(8):1853–1870
5. Guthrie JP (2009) A blind challenge for computational solvation free energies: introduction and overview. J Phys Chem B 113(14): 4501–4507
6. 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
7. 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
8. Dunbar JB, Smith RD, Damm-Ganamet KL, Ahmed A, Esposito EX, Delproposto J, Chinnaswamy K, Kang Y-N, Kubish G, Gestwicki JE (2013) CSAR dataset release 2012: ligands, affinities, complexes, and docking decoys. J Chem Inf Model 53(8): 1842–1852
9. Kryshtafovych A, Monastyrskyy B, Fidelis K (2013) CASP prediction center infrastructure and evaluation measures in CASP10 and CASP ROLL. Proteins Struct Funct Bioinf 82(S2): 7–13
10. Nielsen JE, Gunner M, García-Moreno E (2011) The pKa cooperative: a collaborative effort to advance structure-based calculations of pKa values and electrostatic effects in proteins. Proteins Struct Funct Bioinf 79(12):3249–3259
11. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS (2008) Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. J Med Chem 51(4):769–779
12. Skillman AG, Geballe MT, Nicholls A (2010) SAMPL2 challenge: prediction of solvation energies and tautomer ratios. J Comput Aided Mol Des 24(4):257–258
13. 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
14. Mobley DL, Wymer KL, Lim NM (2014) Blind prediction of solvation free energies from the SAMPL4 challenge. J Comput Aided Mol Des. doi:10.1007/s10822-014-9718-2
15. Mobley DL, Liu S, Lim NM, Wymer KL, Perryman AL, Forli S, Deng N, Su J, Branson K, Olson A (2014) Blind prediction of HIV integrase binding from the SAMPL4 challenge. J Comput Aided Mol Des. doi:10.1007/s10822-014-9723-5
16. Chen W, Chang C-E, Gilson MK (2004) Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. Biophys J 87(5):3035–3049
17. Chang C-E, 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
18. Rekharsky MV, Mori T, Yang C, Ko YH, Selvapalam N, Kim H, Sobransingh D, Kaifer AE, Liu S, Isaacs L (2007) A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy–entropy compensation. Proc Natl Acad Sci USA 104(52):20737–20742
19. 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
20. Nguyen CN, Young TK, Gilson MK (2012) Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. J Chem Phys 137:044101
21. Biedermann F, Uzunova VD, Scherman OA, Nau WM, De Simone A (2012) Release of high-energy water as an essential driving force for the high-affinity binding of cucurbit[n]urils. J Am Chem Soc 134(37):15318–15323
22. Liu S, 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
23. Gibb CL, Gibb BC (2004) Well-defined, organic nanoenvironments in water: the hydrophobic effect drives a capsular assembly. J Am Chem Soc 126(37):11408–11409
24. Cao L, Isaacs L (2014) Absolute and relative binding affinity of cucurbit[7]uril towards a series of cationic guests. Supramol Chem. doi:10.1080/10610278.2013.852674
25. Gibb CL, Gibb BC (2014) Binding of cyclic carboxylates to octa-acid deep-cavity cavitand. J Comput Aided Mol Des. doi:10.1007/s10822-013-9690-2
26. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174
27. Jakalian A, Bush BL, Jack DB, Bayly CI (2000) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. J Comput Chem 21(2):132–146
28. Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem 23(16):1623–1641
29. Zheng L, Yang W (2012) Practically efficient and robust free energy calculations: double-integration orthogonal space tempering. J Chem Theory Comput 8(3):810–823
30. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926
31. Sun Y, Kollman PA (1995) Hydrophobic solvation of methane and nonbond parameters of the TIP3P water model. J Comput Chem 16(9):1164–1169
32. Bennett CH (1976) Efficient estimation of free energy differences from Monte Carlo data. J Comput Phys 22(2):245–268
33. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA Jr (2010) Current status of the AMOEBA polarizable force field. J Phys Chem B 114(8):2549–2564

34. Ren P, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. J Phys Chem B 107(24):5933–5947
35. Grimme S (2012) Supramolecular binding thermodynamics by dispersion-corrected density functional theory. Chem Eur J 18(32):9955–9964
36. Muddana H, Yin J, Sapra N, Fenley A, Gilson M (2014) Blind prediction of SAMPL4 cucurbit[7]uril binding affinities with the mining minima method. J Comput Aided Mol Des. doi:10.1007/s10822-014-9726-2
37. 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
38. Hogues H, Sulea T, Purisima E (2014) Exhaustive docking and solvated interaction energy scoring: lessons learned from the SAMPL4 challenge. J Comput Aided Mol Des. doi:10.1007/s10822-014-9715-5
39. Naïm M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, Siddiqi I, Drabik P, Sulea T, Bayly CI, Jakalian A (2007) Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. J Chem Inf Model 47(1):122–133
40. Corbeil CR, Sulea T, Purisima EO (2010) Rapid prediction of solvation free energy. 2. The first-shell hydration (FiSH) continuum model. J Chem Theory Comput 6(5):1622–1637
41. Mikulskis P, Cioloboc D, Andrejić M, Khare S, Brorsson J, Genheden S, Mata RA, Söderhjelm P, Ryde U (2014) Free-energy perturbation and quantum mechanical study of SAMPL4 octa-acid host-guest binding energies. J Comput Aided Mol Des. doi:10.1007/s10822-014-9739-x
42. Gallicchio E, Lapelosa M, Levy RM (2010) Binding energy distribution analysis method (BEDAM) for estimation of protein—ligand binding affinities. J Chem Theory Comput 6(9):2961–2977
43. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236
44. Gallicchio E, Paris K, Levy RM (2009) The AGBNP2 implicit solvation model. J Chem Theory Comput 5(9):2544–2564
45. Coleman RG, Sterling T, Weiss DR (2014) SAMPL4 & DOCK 3.7: lessons for automated docking procedures. J Comput Aided Mol Des. doi:10.1007/s10822-014-9722-6
46. Kuntz I, Chen K, Sharp K, Kollman P (1999) The maximal affinity of ligands. Proc Natl Acad Sci USA 96(18):9997–10002
47. Rogers KE, Ortiz-Sánchez JM, Baron R, Fajer M, de Oliveira CsAF, McCammon JA (2012) On the role of dewetting transitions in host–guest binding free energy calculations. J Chem Theory Comput 9(1):46–53
48. 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 138:224504
49. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys 120:9665
50. 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
51. Hsiao Y-W, Söderhjelm P (2014) Prediction of SAMPL4 host-guest binding affinities using funnel metadynamics. J Comput Aided Mol Des. doi:10.1007/s10822-014-9724-4
52. Monroe JI, Shirts MR (2014) Converging free energies of binding in cucurbit[7]uril and octa-acid host-guest systems from SAMPL4 using expanded ensemble simulations. J Comput Aided Mol Des. doi:10.1007/s10822-014-9716-4
53. Coleman RG, Carchia M, Sterling T, Irwin JJ, Shoichet BK (2013) Ligand pose and orientational sampling in molecular docking. PLoS ONE 8(10):e75992
54. Zwanzig RW (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. J Chem Phys 22:1420
55. Kollman P (1993) Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 93(7):2395–2417
56. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13(8):1011–1021
57. Lyubartsev A, Martsinovski A, Shevkunov S, Vorontsov-Velyaminov P (1992) New approach to Monte Carlo calculation of the free energy: method of expanded ensembles. J Chem Phys 96:1776
58. König G, Boresch S (2011) Non-Boltzmann sampling and Bennett's acceptance ratio method: how to profit from bending the rules. J Comput Chem 32(6):1082–1090
59. Hawkins G, Giesen D, Lynch G, Chambers C, Rossi I, Storer J, Li J, Zhu T, Thompson J, Winget P (2004) AMSOL-version 7.1. University of Minnesota, Minneapolis, based in part on AMPAC-version 2
60. Grimme S (2011) Density functional theory with London dispersion corrections. Wiley Interdiscip Rev Comput Mol Sci 1(2):211–228
61. Sure R, Grimme S (2013) Corrected small basis set Hartree-Fock method for large systems. J Comput Chem 34(19):1672–1685
62. Řezáč J, Fanfrlík 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
63. Korth M (2010) Third-generation hydrogen-bonding corrections for semiempirical QM methods and force fields. J Chem Theory Comput 6(12):3808–3816
64. Momany FA, Rone R (1992) Validation of the general purpose QUANTA® 3.2/CHARMm® force field. J Comput Chem 13(7):888–900
65. Gilson MK, Gilson HS, 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
66. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 31(4):671–690
67. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem 97(40):10269–10280
68. Hampel C, Werner HJ (1996) Local treatment of electron correlation in coupled cluster theory. J Chem Phys 104:6286
69. Purisima EO (1998) Fast summation boundary element method for calculating solvation free energies of macromolecules. J Comput Chem 19(13):1494–1504
70. Klamt A (1995) Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. J Phys Chem 99(7):2224–2235
71. Klamt A, Schüürmann G (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2(5):799–805
72. Sitkoff D, Sharp KA, Honig B (1994) Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem 98(7):1978–1988
73. Chang CE, Gilson MK (2003) Tork: conformational analysis method for molecules and complexes. J Comput Chem 24(16):1987–1998

74. Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci USA 99(20):12562–12566
75. Limongelli V, Bonomi M, Parrinello M (2013) Funnel metadynamics as accurate binding free-energy method. Proc Natl Acad Sci USA 110(16):6358–6363
76. 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