Exhaustive docking and solvated interaction energy scoring: lessons learned from the SAMPL4 challenge

Hervé Hogues · Traian Sulea · Enrico O. Purisima

Abstract We continued prospective assessments of the Wilma–solvated interaction energy (SIE) platform for pose prediction, binding affinity prediction, and virtual screening on the challenging SAMPL4 data sets including the HIV-integrase inhibitor and two host–guest systems. New features of the docking algorithm and scoring function are tested here prospectively for the first time. Wilma–SIE provides good correlations with actual binding affinities over a wide range of binding affinities that includes strong binders as in the case of SAMPL4 host–guest systems. Absolute binding affinities are also reproduced with appropriate training of the scoring function on available data sets or from comparative estimation of the change in target's vibrational entropy. Even when binding modes are known, SIE predictions lack correlation with experimental affinities within dynamic ranges below 2 kcal/mol as in the case of HIV-integrase ligands, but they correctly signaled the narrowness of the dynamic range. Using a common protein structure for all ligands can reduce the noise, while incorporating a more sophisticated solvation treatment improves absolute predictions. The HIV-integrase virtual screening data set consists of promiscuous weak binders with relatively high flexibility and thus it falls outside of the applicability domain of the Wilma–SIE docking platform. Despite these difficulties, unbiased docking around three known binding sites of the enzyme resulted in over a third of ligands being docked within 2 Å from their actual poses and over half of the ligands docked in the correct site, leading to better-than-random virtual screening results.

Introduction

Accurate prediction of structural and energetic aspects of binding in aqueous solution is critical for successful structure-based drug design and the understanding of molecular recognition in biological systems. Binding affinity prediction methods range from the relatively slow but thermodynamically rigorous pathway approaches such as free energy perturbation (FEP) and thermodynamic integration (TI) [1, 2], to the faster end-point approaches relying on binding affinity scoring functions that can be classified into three main categories: force-field-based, knowledge-based, and empirical [3–9]. Although many end-point methods are based on implicit solvent descriptions, the solvent potential of the mean force theory ensures that, given adequate configurational sampling, these methods can be as rigorous as alchemical pathway methods based on explicit solvent description [10, 11]. A popular method in the force-field-based group is MM-PB(GB)/SA [12–14], which combines molecular mechanics-based terms with continuum solvation terms.

Solvated interaction energy (SIE) [15–17] is another end-point force-field-based scoring function that approximates binding affinity by an interaction energy contribution and a desolvation free energy contribution, each of them further made up of electrostatic and nonpolar components. Calibrated on a diverse dataset of 99 protein–ligand

H. Hogues · T. Sulea · E. O. Purisima (✉)
Human Health Therapeutics, National Research Council Canada, 6100 Royalmount Avenue, Montreal, QC H4P 2R2, Canada
e-mail: Enrico.Purisima@nrc-cnrc.gc.ca

complexes [15], SIE achieves a reasonable transferability across a wide variety of protein–ligand systems for which it predicts absolute binding affinities within the experimental range as shown by various test cases reported in the literature [17, 18]. External testing of the standard SIE parametrization in the CSAR-2010 dataset of 343 protein–ligand complexes diverse with respect to ligands and targets predicted absolute binding affinities with a mean-unsigned-error (MUE) of about 2 kcal/mol [18].

A docking procedure is required in order to apply SIE in absence of crystallographic ligand poses. To this end we developed Wilma, an exhaustive docking program that has the required efficiency for large-scale virtual screening of small-molecule libraries. Owing to its exhaustive nature as well as to its fast empirical pose-ranking function calibrated on crystal structures of protein–ligand complexes, the top-ranked pose produced by Wilma has been shown to be consistently close to the experimental pose for drug-like ligands.

Both SIE and Wilma have been employed for blind testing in previous editions of Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) organized by OpenEye, Inc. A reasonable performance of SIE in binding affinity prediction for the SAMPL1 set of kinase inhibitors with available cognate crystal structures had been noted [19]. In SAMPL3, the Wilma–SIE virtual screening platform achieved good enrichment of true positives from a dataset of fragment-size ligands against trypsin, with an AUC of about 0.7 for a receiver-operating-characteristic (ROC) curve characterized by an excellent early enrichment performance [20]. Binding affinity predictions for trypsin–ligand and host–guest complexes in SAMPL3 were generally within 2 kcal/mol of the experimental values but rank ordering of affinities within 2 kcal/mol was not well predicted.

In this paper, we continue prospective testing of the Wilma–SIE docking–scoring platform for both virtual screening and binding affinity predictions. We tested our methods on both molecular systems proposed in SAMPL4. One the one hand were the two relatively small hosts with their surprisingly high-affinity guest ligands. On the other was the much larger homodimeric HIV-1 integrase that can bind various inhibitors at several sites with much weaker affinities than one would expect based on the shape of the enzyme pockets and the size of the ligands. Several methodological and system-dependent properties are explored in this study. These include: (1) a new virtual screening scoring function replacing the surface-based terms with a penalty term for non-complementary polar and non-polar interactions, (2) the role of vibrational entropy change and symmetry corrections to the absolute magnitude and correlation of binding affinity predictions, respectively, studied on the host–guest systems, (3) the effect of the size of the sampled docking space for docking and enrichment of actives, studied on HIV-integrase pose prediction and virtual screening datasets, (4) a more advanced continuum solvation model and the use of a common structure of the target for binding free energy predictions, studied on the HIV-integrase affinity dataset.

Methods

Wilma docking

The docking software Wilma uses a brute-force searching approach where the interaction with the rigid protein of all the discrete rotational and translational states of every ligand conformation generated by Omega (OpenEye Scientific Software, New Mexico) is examined. Using an efficient filtering method, the program exhaustively enumerates, scores and ranks all the ligand poses that do not overlap with the protein. Docking is done within one or several predefined rectangular volumes with a translation step size of 0.5 Å. The discrete rotation of the ligand is adjusted to insure that the maximum movement of any atom between adjacent orientations is less than 1 Å. The ligand conformations generated by Omega are controlled by setting the internal energy cutoff to 20 kcal/mol and adjusting the RMSD clustering parameter to produce at most 5,000 conformations.

The original weighted 5-term scoring function used for Wilma docking was trained to recover the most native states using 320 protein–ligand complexes from the curated CSAR dataset [21]

$$\text{WilmaScore1} = w_1 E_{\text{coul}} + w_2 E_{\text{vdw}} + w_3 E_{\text{HB}} + w_4 E_{\text{psc}} + w_5 E_{\text{npsc}}$$

(1)

This scoring function includes a coulombic interaction term, E_{coul}, a van der Walls 6-12 Lennard-Jones potential, E_{vdw}, an explicit H-bond term, E_{HB}, which considers donor and acceptor orientations, and two surface (polar and non-polar) complementarity terms, E_{psc} and E_{npsc}. For this study we calibrated a different version of the scoring function for Wilma docking that replaces the two surface complementary terms by a term, E_{flaws}, which introduces an energetic penalty for flaws present in the docked pose in terms of protein–ligand polar complementarity.

$$\text{WilmaScore2} = w_1 E_{\text{coul}} + w_2 E_{\text{vdw}} + w_3 E_{\text{HB}} + w_4 E_{\text{flaws}}$$

(2)

These flaws account for the obstruction of polar groups by non-polar or like-charged polar groups. Introduction of the E_{flaws} model is an attempt to reduce occasional top-ranked

poses and false-positive ligands that are "flawed" due to the presence of buried partially charged atoms without formation of electrostatically complementary interactions in the bound state, which were still observed when using the surface complementarity terms. This empirical geometrical model poses a more stringent electrostatic desolvation penalty on such unfavorable interactions (flaws) in addition to addressing the charge sign of polar interactions in comparison with the former surface-based model. Further description and implementation details of the E_{flaws} term are provided in the Supplementary Material. The Wilma scoring function was used exclusively for structure prediction, i.e., to select the top-ranked docked pose.

Solvated interaction energy (SIE) calculations

Scoring of binding affinities was carried out using the SIE end-point force-field based method [15–18], which approximates the binding free energy from the electrostatic and non-polar components of the interaction energy and the desolvation free energy

$$\text{SIE} = \alpha(E_{coul} + E_{vdw} + E_{RF(BEM)} + E_{npsolv}) + C \quad (3)$$

where E_{coul} and E_{vdw} describe solute–solute interactions by intermolecular coulombic and van der Waals interaction energies in the bound state calculated with AMBER and GAFF molecular mechanics force fields [22–24]. Desolvation effects are described by $E_{RF(BEM)}$, the change in the reaction field energy between the bound and free states calculated with a continuum model based on a boundary element solution to the Poisson equation using the BRI BEM program [25, 26] and a solute dielectric constant $D_{in} = 2.25$, and E_{npsolv}, the non-polar desolvation approximated from a linear proportionality with the change in solute molecular surface area [27–29]. The free state of the system is obtained by rigid separation of the interacting molecules from the bound state. Partial atomic charges for protein atoms are taken from the AMBER force field, whereas organic solutes are assigned AM1-BCC partial charges [30, 31]. α is a global scaling factor of the total raw solvated interaction energy relating to the scaling of the binding free energy due to configurational entropy effects [32, 33]. The standard parameters of the SIE function in Eq. (3) are $\alpha = 0.1048$ and $C = -2.89$ kcal/mol calibrated against a protein–ligand training dataset of 99 complexes refined by restrained energy minimization [15].

We also explored prospectively a different SIE function in which the solvation terms are replaced by our latest continuum solvation model FiSH that captures some of the properties of the first shell of hydration [34, 35]. For example, the electrostatic desolvation in the FiSH model, $E_{RF(FISH)}$, can account for charge asymmetry effects. Also,

instead of a single surface-area-based term for all non-electrostatic component of solvation, FiSH includes an additional continuum van der Waals term, E_{cvdw}, to more accurately describe the solute–solvent non-polar interactions, and a separate surface-area based cavity term, E_{cav}. Unlike the default solvation model within SIE, which uses a solute dielectric of 2.25, the FiSH model uses a solute dielectric of 1.0. The modified SIE + FiSH scoring function then has the form

$$\text{SIE} + \text{FiSH} = \alpha(E_{coul} + E_{vdw} + E_{RF(FISH)} + E_{cvdw} + E_{cav}) + C$$

$$(4)$$

where the parameters $\alpha = 0.1232$ and $C = 1.46$ were obtained by training against the same 99 protein–ligand data set used for the original SIE function [15].

Finally, another SIE variant that implements two terms from the Wilma docking program, the explicit hydrogen bonding term, E_{HB}, and the energetic penalty term for flaws of protein–ligand complementarity, E_{flaws},

$$\text{SIE} + \text{HB} + \text{FLAW} = \alpha(E_{coul} + E_{vdw} + E_{RF(BEM)} + E_{npsolv}) + \beta E_{HB} + \delta E_{flaws} + C$$

$$(5)$$

were calibrated against the 320 protein–ligand complexes from the curated CSAR dataset [21], leading to weighting factors $\beta = -0.4$ and $\delta = 1.2$, while keeping α and C at the default values in Eq. (3).

Prior to SIE, SIE + FiSH and SIE + HB + FLAW calculations, all complexes were refined by constrained energy minimization as described previously [18, 20].

The average CPU time for a Wilma–SIE calculation was of the order of 10 min for a typical protein–ligand complex in the HIV integrase virtual screening exercise. It generally took Wilma about 0.1 s to exhaustively dock one conformation of a ligand. For each ligand up to 5,000 Omega-generated conformations were docked. The docked poses were then clustered and representatives from each cluster were rescored with SIE. Each protein–ligand representative took about 20 s to rescore.

Structural preparation

HIV-integrase data set for virtual screening

The 1.9-Å-resolution crystal structure of the homodimeric HIV-1 integrase catalytic core domain prepared for virtual screening was taken from the PDB entry 3NF8 as suggested by the SAMPL4 organizers. Structural preparation of the dimeric structure was done in SYBYL 8.1.1 (Tripos, Inc., St. Louis, MO). The crystallographic water molecules, acetate and sulfate ions, and co-solvent and ligand

molecules were removed. Chain termini of the dimeric structure, including those arising from the disordered loop between residues Lys188 and Gly193 were capped with acetyl and methylaminyl groups. Hydrogen atoms were added, with the protonation states of most ionizable side-chains assigned for neutral pH. Exceptions include the side chains of Asp64 and Glu92 in both monomers, which were protonated. Tautomeric and protonation states of His residues were manually assigned after visual inspection in order to maximize the H-bonding network, noting the protonated forms assigned to His72 and His183 in both monomers. Polar hydrogen atoms were oriented to maximize H-bonding and then the structure was refined by energy minimization with the AMBER force field using harmonic constraints of 3 and 20 kcal/(molÅ^2) for the non-hydrogen side-chain and backbone atoms, respectively.

In order to prepare the database of ligands for virtual screening, we first verified the provided protonation states at neutral pH and introduced alternate states for 13 ligands. These include deprotonation of pyridine N atoms in ligands AVX101125_0, AVX17228_0 and AVX17231_0, deprotonation of aromatic amine N atoms in AVX17264_0, AVX17264_1, AVX38752_0, AVX38752_1, AVX40869_0, AVX40872_0 and AVX62526_0, deprotonation of the imidazole ring in ligand AVX62778_0, and tautomerism between piperidine N atoms in ligands AVX40989_0 and AVX40989_1. Partial charges were calculated with the AM1-BCC method [30, 31], as implemented in Molcharge (Open-Eye, Inc.), using as input the lowest-energy conformation generated by Omega (OpenEye, Inc.).

HIV-integrase data set for affinity prediction

Two sets of structures were prepared. In one set cognate protein structures for each ligand were used as inferred from the corresponding crystal structures. In the other set a common protein structure was used for all ligands. In the cognate set, the eight crystal structures of complexes provided for prospective predictions as well as the suggested control structure 3ZSQ of a complex with previously measured binding affinity were prepared in a similar way as described earlier for the virtual screening data set, followed by constrained energy-minimization of the complexes around the ligands as required for SIE calculations [18, 20]. The number of protein atoms was kept the same in all these complexes, which required deletion of C-terminal Ala residue in one of the structures (the AVX17557 complex). In the common set, the cognate protein structures in all these complexes were replaced by the 3NF8 structure prepared for virtual screening, after root-mean-square fitting of backbone atoms, and then refined by the same energy minimization protocol.

Host–guest data sets

The provided structure of the cyclic cucurbit[7]uril (CB7) and OctaAcid hosts used for Wilma docking and SIE affinity scoring were first energy-minimized with the GAFF force-field, AM1-BCC partial charges and a distance-dependent dielectric constant. The cyclic OctaAcid host, containing eight carboxylate side chains, was used in the state corresponding to the formal net charge of $-8e$. The rotameric states of its four aliphatic carboxylate side-chains were manually changed into a symmetrical geometry prior to energy minimization. The structures of the all guests (14 amines as CB7 guests and 9 carboxylic acids as OctaAcid guests) were docked in the most probable protonation states at the corresponding experimental pHs (as provided, with the exception of the CB7 guest #10 for which an alternate state corresponding to a mono-protonated piperidine ring was also docked). A training set of seven guests with measured binding affinities for the CB7 host [36] was prepared as described previously [20].

Vibrational entropy calculations

The relatively small size of the host–guest systems allows the direct application of a normal mode analysis (NMA) to compute the vibrational entropy change upon binding [37]. Here, the AMBER force field with a distance-dependent dielectric was used for the minimization and construction of the mass-weighted Hessian matrix.

Results and discussion

Host–guest affinity prediction

We submitted two prospective models for each of the CB7 and OctaAcid host–guest affinity prediction data sets, one based on the standard SIE scoring function in Eq. (3) and the other one on the SIE + HB + FLAW function in Eq. (5). We used our exhaustive docking program Wilma to arrive at bound conformations for host–guest complexes. The search space was defined large enough to allow docking of the guest at any contact position around the host. In general, the top-scored pose for all guests was found to bind through the central hole-region of the hosts. Both hosts are macrocycles having a circular geometry with a central hole where certain guests are recognized with surprisingly high affinity given the relatively small size of these systems [36]. Whereas CB7 is a neutral host, OctaAcid is negatively charged due to eight carboxylate side chains disposed peripherally and away from the macrocycle [38, 39]. Their guests are depicted in Figures S1 and S2 from the Supplementary Material.

Table 1 Performance of host–guest and HIV-integrase binding affinity predictions

System	ID#[a]	Method[b]	RMSE	MUE	Kτ	R	Slope
CB7	187	SIE	3.90	3.46	0.54	0.74	0.18
CB7	188	SIE + HB + FLAW	3.92	3.48	0.57	0.75	0.19
CB7	Retro	rSIE	1.67	1.24	0.54	0.74	0.62
CB7	Retro	rSIE, symmetry correction	1.72	1.29	0.56	0.74	0.60
CB7	Null		2.16	1.86	n/a	n/a	n/a
OctaAcid	185	SIE	1.32	1.19	0.70	0.81	0.24
OctaAcid	186	SIE + HB + FLAW	1.33	1.19	0.70	0.87	0.28
OctaAcid	Retro	rSIE	1.59	1.39	0.70	0.81	0.82
OctaAcid	Retro	rSIE, symmetry correction	1.50	1.23	0.59	0.76	0.81
OctaAcid	Null		1.24	1.00	n/a	n/a	n/a
HIV-IN	182	SIE, cognate structures	3.92	3.85	−0.36	−0.67	−0.76
HIV-IN	183	SIE, common structure	3.34	3.25	−0.64	−0.78	−0.80
HIV-IN	184	SIE + FiSH, cognate structures	1.84	1.33	−0.14	−0.35	−1.03
HIV-IN	Retro	SIE + FiSH, common structure	1.21	0.90	−0.50	−0.55	−1.30
HIV-IN	Null		0.40	0.34	n/a	n/a	n/a

RMSE: root-mean-square error (kcal/mol); MUE: mean unsigned error (kcal/mol); Kτ: Kendall tau; R: correlation coefficient; slope: correlation slope. Note that the RMSE and MUE values reported for the host–guest systems are the absolute values and hence are larger than the corresponding values reported in the SAMPL4 overview paper by Muddana et al. [46] which are values for predictions of relative binding affinities by taking one of the host–guest systems as reference

[a] Submission number. Retro: retrospective model. Null: predicted affinities set to the experimental mean

[b] SIE: $\alpha = 0.1048$ and $C = -2.89$ in Eq. (3); rSIE: rescaled SIE function based on an external training data set for CB7 with $\alpha = 0.3572$ and $C = 2.24$ in Eq. (3); SIE + FiSH: $\alpha = 0.1232$ and $C = 1.46$ in Eq. (4); SIE + HB + FLAW: $\alpha = 0.1048$, $\beta = -0.4$, $\delta = 1.2$ and $C = -2.89$ in Eq. (5)

The statistical performances of the models are listed in Table 1 (see Table S1 for the values of the predicted binding affinities). Since the results with the two scoring functions were similar we will discuss only those based on the SIE function. We see that there is good correlation with SIE for both hosts but the slopes are small, that is, predicted range much smaller than the experimental one. One way to modulate the correlation slope is to rescale the SIE function in terms of the enthalpy–entropy compensation factor α in Eq. (3) specifically for the system being investigated. This is justified since is has been previously shown that the CB7 host, for example, requires a higher energy efficiency factor, that is, the degree to which attractive forces are effective in generating binding free energy, rather than being cancelled by entropy losses, than the β-cyclodextrin host [33, 40, 41]. This points towards a larger value for the α scaling factor in the SIE formulation. Hence, we explored this possibility retrospectively by deriving a rescaled SIE function based on a previously published data for guests binding to CB7 (7 complexes) [36]. This training model leads to an α scaling factor of 0.3572 (with a positive constant $C = 2.24$), hence a significantly larger scaling than that for the standard SIE function (0.1048), in agreement with previous observations [40, 41]. Application of the rescaled SIE function to the SAMPL4 CB7 data set led to a retrospective model with a much-improved correlation slope (Table 1; Fig. 1a, b).

The weaker entropy compensation in CB7 binding as compared to proteins is likely due to the rigidity of the CB7 host to begin with, resulting in a reduced loss of entropy upon complex formation. As a way to support this explanation, we noted that the slope for the cyclic CB7 host is not only larger than that for proteins but also larger than that obtained for the acyclic host analog to CB7 we examined in SAMPL3 [20]. In order to assess qualitatively the reduced entropic costs of binding to the cyclic CB7 host, we have designed a computational experiment comparing the cyclic CB7 and its acyclic cucurbituril analog studied (as host-1) in SAMPL3 [20] (Fig. 2). The loss of vibrational entropy of the target host upon guest binding (a cyclohexyl diamine) was calculated by normal mode analysis in each system. After adding the loss of rotational and translational entropy of the ligand upon binding we calculated entropic losses $-T\Delta S_{binding}$ of only +9.7 kcal/mol in the case of the cyclic CB7 host compared with the larger loss of +13.3 kcal/mol for the acyclic host. This corroborates nicely the smaller enthalpy–entropy compensation and hence a larger scaling factor of binding free energy for the cyclic analog relative to the acyclic one. Furthermore, since OctaAcid is also a cyclic host, we applied the scaling
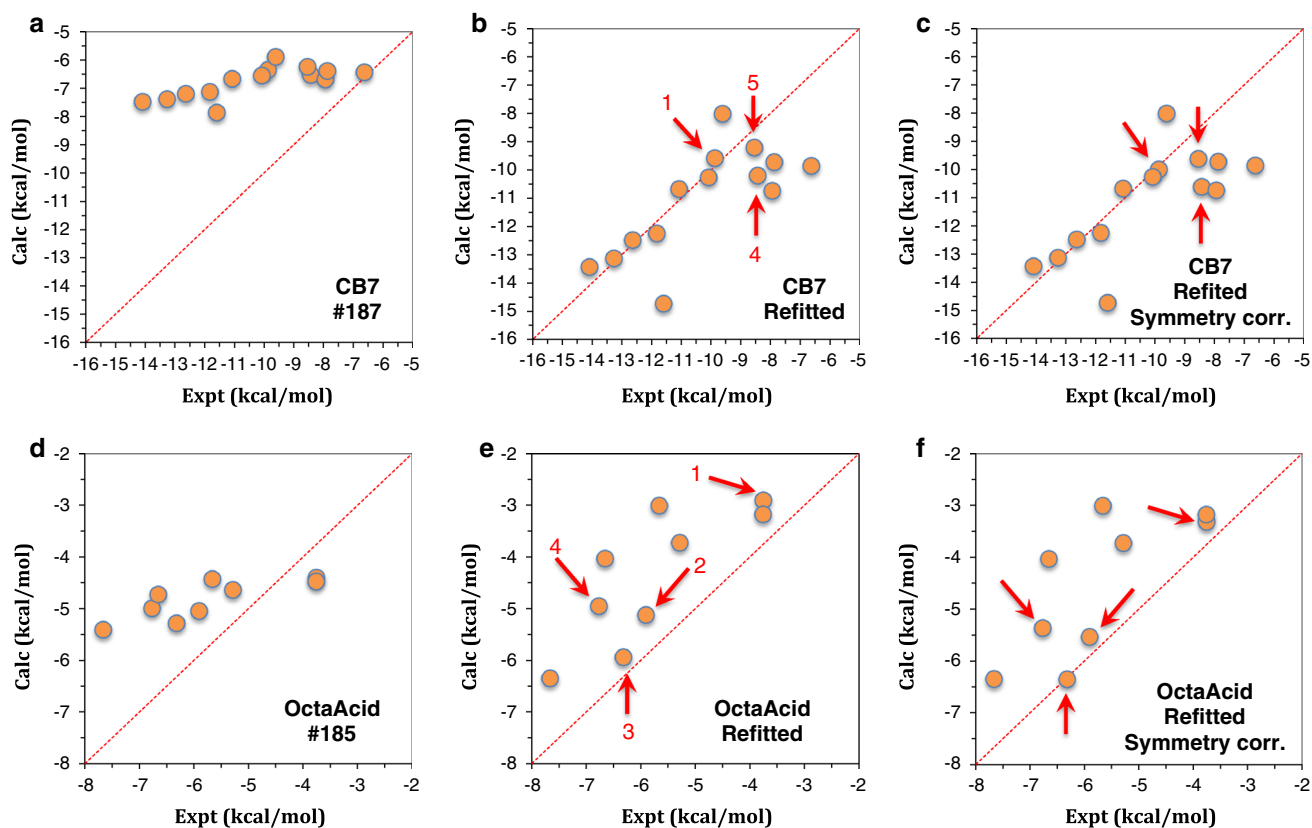
Fig. 1 Scatter plots of calculated versus actual binding affinities for the host–guest systems and various prediction models. **a** CB7 host, standard SIE, submission #187; **b** CB7 host, refitted SIE (based on external published data for 7 complexes of CB7), retrospective; **c** CB7 host, refitted SIE, symmetry correction, retrospective; **d** OctaAcid host, standard SIE, submission #185; **e** OctaAcid host, refitted SIE (based on external published data for 7 complexes of CB7), retrospective; **f** OctaAcid host, refitted SIE, symmetry correction, retrospective. The *diagonal dashed line* indicates a perfect correlation

Fig. 2 Role of vibrational entropy of the target to free energy scaling studied by comparing the cyclic host CB7 with an acyclic host analog. The ligand bound to each host is shown in ball-and-stick representation

Cyclic host (CB7)   Acyclic host
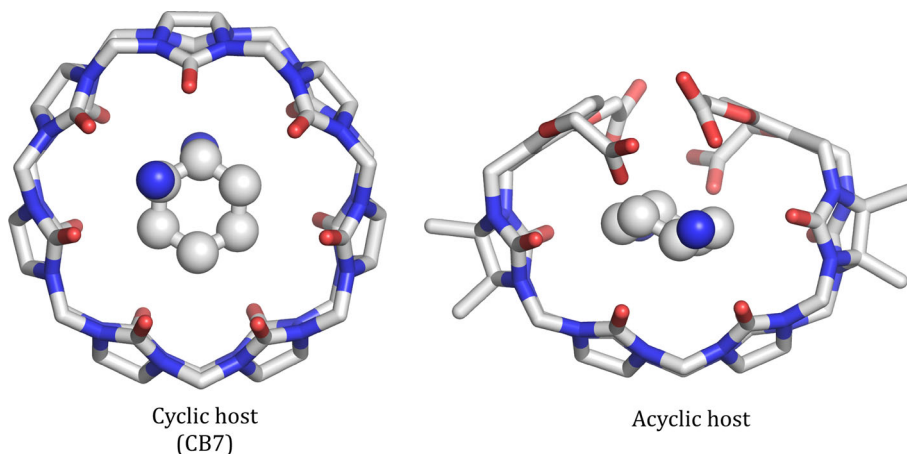
factor derived for CB7 and significantly improved the correlation slope for this system as well (Fig. 1d, e). All these data suggest that entropic scaling of free energy is system-dependent and can be calibrated if data is available for the system under investigation. If not enough data with a good dynamics range is available, vibrational entropy calculations by normal mode analysis may provide an

alternative for comparison between various systems and appropriate adjustment of the scaling coefficient.

The symmetry of the hosts and some of the guests can have consequences on binding free energies [42, 43]. Since the host symmetry affects all ligands equally, only the guest (ligand) symmetry corrections need to be considered for relative binding free energy calculations. These

corrections (~0.4 kcal/mol for a twofold symmetry) applied to the retrained SIE scoring function for both CB7 and OctaAcid systems have a marginal effect (Fig. 1c, f).

HIV integrase affinity prediction

We submitted SIE and SIE + FiSH prospective predictions for the HIV-integrase binding affinity data set (Table 1, Table S1), which consists of eight inhibitors (depicted in Figure S3) against the binding site for the LEDGF/p75 cellular cofactor of HIV-1 integrase (termed the LEDGF site throughout the rest of the paper). Previously unreleased crystal structures of these enzyme-inhibitor complexes were made available for this blind challenge. We first used these cognate protein structures for generating SIE predictions of binding free energies (submission #182). The correlation between these predictions and the actual values is quite poor (Table 1). It is worth noting that the dynamic range of binding affinities in this data set is extremely narrow (1.2 kcal/mol), so from this viewpoint the SIE blind prediction was successful in that the dynamic range of predicted binding affinities was similarly narrow (1.4 kcal/mol). However, SIE was trained and externally tested to achieve a performance of about 2 kcal/mol mean-unsigned error [15, 18, 20] and hence it is not capable of

reliably ranking binding affinities within smaller dynamic ranges. The absolute magnitude of binding affinities was also overestimated by SIE in this data set (Fig. 3a). This may relate to the fact that SIE suffers from a certain mass bias and the ligands in this data set are relatively large (with molecular weights between 364 and 574 Da) for their measured binding affinities (0.2–1.5 mM dissociation constants).

We then wanted to test whether small structural changes afforded in the protein target by various cognate crystal structures are contributing favorably or not to SIE predictions. We refer here to changes that are distributed all around the protein molecule and involve main-chain and side-chain fluctuations that are not necessarily limited to transitions over large torsional barriers. We also noted that several exposed side chains close to moieties that are common for all these ligands, for example Gln95 and His171, experience significantly different rotameric states in different crystal structures. Therefore, we replaced the cognate protein structures with a single external structure (taken from the available PDB structure 3NF8). The SIE prospective predictions from this common target structure experiment (submission #183) did not worsen the predictions, which are still within a narrow (1.1 kcal/mol) dynamic range, and actually we noticed a slight

Fig. 3 Scatter plot of predicted versus experimental binding affinities of HIV-integrase ligands. a SIE function on cognate structures (submission #182); b SIE function on single structure (submission #183); c SIE function on cognate structures (submission #184); d SIE + FiSH function on single structure (retrospective). The diagonal dashed line indicates a perfect correlation
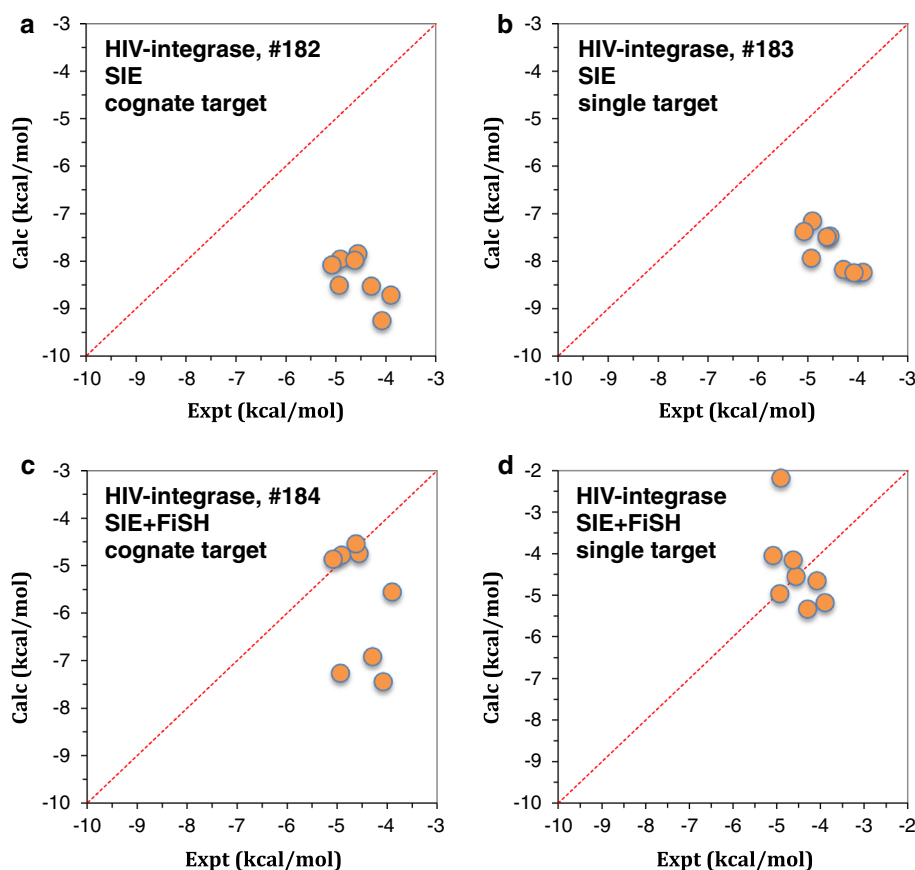
improvement in the magnitude of absolute predictions (Fig. 3b). This indicates that the common protein structure is a good strategy for noise reduction in predictions with the SIE and related methods, which are based on a single conformation of the complex. It also more faithfully represents the routine application of SIE in most virtual screening campaigns. If small conformational movements are needed close to the ligand, then those can be introduced on the common scaffold thus eliminating the noise introduced by distant movements.

Application of the SIE + FiSH variant of the scoring function improved the absolute magnitudes of predictions for half of the complexes with the cognate (multiple) target structure approach (submission #184, Fig. 3c) and for all but one complex in the common (single) target structure approach (retrospective prediction, Fig. 3d). This indicates that SIE + FiSH may be less sensitive to size bias than SIE, reinforcing some of the earlier findings based on our experience in SAMPL3 [20]. It is also apparent from the current results that the spread of the predictions with SIE + FiSH is larger than the SIE-based spread and the experiment, which indicates that this model is more sensitive to protein structural changes. This further reinforces the value of using the common structure approach.

HIV-integrase pose prediction and virtual screening

The HIV-integrase virtual screening challenge consisted of identifying a set of binders from a final full set of 305 molecules, some of which are stereoisomers of the same compounds. Retrospectively, there were 56 distinct binders in this data set, the rest consisting of proven non-binding decoy molecules that are structurally similar analogs of the binders. One peculiarity of this virtual screening challenge was that the target, HIV-integrase, can bind ligands at three distinct sites (actually six considering the dimer): the LEDGF site, the Y3 site, and the fragment site, although most binders included in this set bind to the LEDGF site [44, 45]. We directed virtual screening of the full set to all three sites and selected the best scoring pose overall using the standard SIE scoring function in Eq. (3) (submission #146) and the SIE variant that includes hydrogen bonding and flaws terms as in Eq. (5) (submission #147). In a third submission (#148) we also used the newer SIE + HB + FLAW function but ranked compounds based on the scores calculated at the LEDGF site only.

Obviously, the success or failure of the virtual screening experiment hinges greatly on the docking step. Hence, before discussing the virtual screening results, we wanted to get a feel for the docking accuracy based on the SAMPL4 pose prediction challenge consisting in the HIV-integrase binders together with their known actual poses. Hence, our pose prediction submissions (#154, #155 and

#156) were essentially those from our corresponding virtual screening runs mentioned above (submission #146, #147 and #148, respectively). An overview of our pose prediction results over all binders is shown in Table 2. About a third of ligands were docked well (up to 2 Å RMSD from the actual pose) by Wilma when pose selection was done with the standard SIE function over all three sites. Slightly less ligands were docked well when scored by SIE + HB + FLAW overall all three sites and also when docking was directed only around the LEDGF site. Half of the ligands were docked closer than 4.52 Å RMSD from the actual pose with standard SIE scoring over all three sites, which is a reasonable performance.

The interpretation of the pose prediction challenge in SAMPL4 is complicated by the existence of several binding sites for various ligands as well as multiple binding of some of the compounds at more than one site. In the same time, this also represents a very stringent test of the docking method. Half of the ligands are docked at the correct site using the standard SIE scoring, with an average RMSD over this fraction of ligands of only 2.36 Å and with half of this fraction of ligands docked closer than 1.60 Å to the actual pose (Table 2). When docking was constrained only around the LEDGF site, 85 % of the ligands were docked in the correct site, because the rest of the compounds actually bind to other sites. The important metrics to remember in this case are the average and median RMSD values of 6.96 and 5.20 Å calculated for this fraction of compounds. These relatively large RMSD values indicate a certain level of misdocking within the relatively wide docking region that we set up around the LEDGF site.

To gain more insight into the performance of our docking program in this system, we focused on a subset of eight compounds from HIV-integrase affinity prediction challenge presented earlier for which we had access to the actual poses. These compounds bind to the same pocket in the LEDGF site. However, although these compounds were docked in the box around the LEDGF site, Wilma docking positioned correctly (RMSD < 2 Å) only one of the eight compounds. There are several compelling reasons for this poor docking result.

First, the box defining the docking space around the LEDGF site was much larger than the actual binding site of the ligands. We purposely defined a larger region because of the presence of a deep pocket adjacent to the actual binding site of these ligands (Fig. 4a). We found that most compounds docked into the deeper pocket rather than in the much shallower actual pocket. This was irrespective of whether SIE or SIE + HB + FLAW scoring functions were used to rank the poses generated by the Wilma docking program. Retrospectively repeating the docking on a smaller box focusing strictly around the actual binding

Table 2 Performance of pose predictions for binders of HIV-integrase

ID#[a]	Method[b]	Fraction ≤ 2Å RMSD (%)	Average RMSD (Å)	Median RMSD (Å)	In the correct site		
					Fraction (%)	Average RMSD (Å)	Median RMSD (Å)
154	SIE, three sites	38	12.51	4.52	53	2.36	1.60
155	SIE + HB + FLAW, three sites	27	15.65	18.12	43	3.83	1.77
156	SIE + HB + FLAW, LEDGF site—*wide*	30	8.00	6.18	85	6.96	5.20

[a] Submission number

[b] SIE: Eq. (3); SIE + HB + FLAW: Eq. (5); three sites: docking was performed in three distinct regions around the Y3 site, the fragment site and the LEDGF site; LEDGF site—*wide*: docking was performed in a large area around the LEDGF site
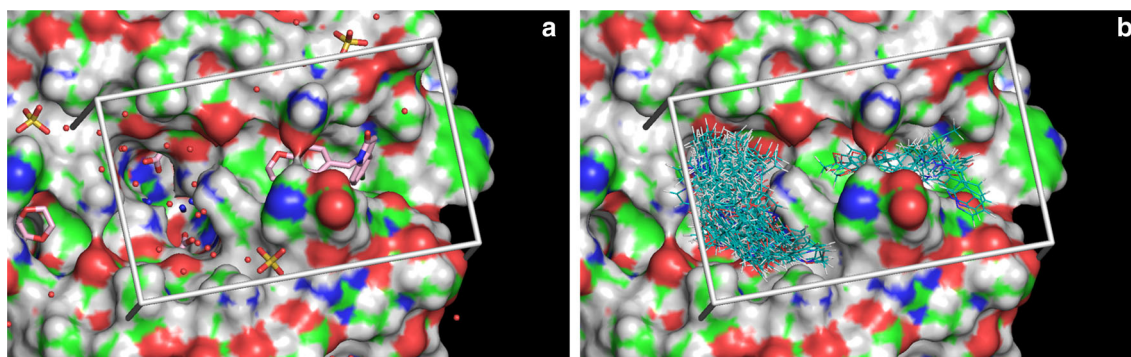
Fig. 4 *Box* defining the region used for virtual docking. HIV-integrase is represented as a molecular surface. **a** The *box* extends beyond the actual binding location of the ligand and include an adjacent deep pocket filled with structured water molecules (*red spheres*) and an acetate ion. **b** Low-energy docked poses are found mostly in the adjacent deep pocket and less in the shallower actual binding site

location resulted in correct docking of five out of the eight ligands. Overall, these results may indicate that favorable van der Waals interactions in the deep pocket are overwhelming the cost of displacing water and ions from this location, leading to incorrect pose ranking. Hence, some further improvement of our scoring functions is warranted.

Secondly, as already mentioned, these HIV-integrase inhibitors are weak binders (>100 μM dissociation constant). They are also highly flexible, having more than eight rotatable bonds. Weak and flexible ligands are promiscuous to bind at several locations at the protein surface. This is reflected in our docking results, where we find that poses in the wrong pocket outscored the correct poses by just 0.2 kcal/mol (Fig. 4b). Docking of flexible weak-binding ligands is highly prone to generate false positives, i.e., good-scoring incorrect pose. Handling highly flexible ligands is also difficult because the bioactive conformation may not be readily generated before docking (although Omega failed to generate the bioactive conformation for only one of the eight ligands). Hence, this challenge seems to fall outside of the applicability domain of our Wilma–SIE docking–scoring virtual screening platform, which is designed to reliably differentiate not-too-flexible (less than eight rotatable bonds) strong binders (at least sub-μM) from non-binders.

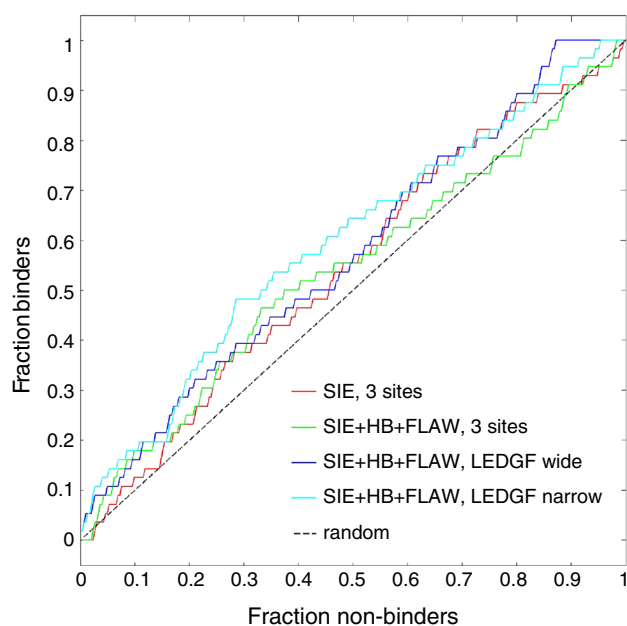**Table 3** Performance of virtual screening against HIV-integrase

ID#[a]	Method[b]	EF (10 %)	EF (20 %)	AUC (ROC)
146	SIE, three sites	1.25	1.16	0.55
147	SIE + HB + FLAW, three sites	1.79	1.25	0.54
148	SIE + HB + FLAW, LEDGF site—*wide*	1.61	1.52	0.58
Retro	SIE + HB + FLAW, LEDGF site—*narrow*	1.79	1.61	0.60

[a] Submission number. Retro: retrospective prediction

[b] SIE: Eq. (3); SIE + HB + FLAW: Eq. (5); three sites: docking was performed in three distinct regions around the Y3 site, the fragment site and the LEDGF site; LEDGF site—*wide*: docking was performed in a large area around the LEDGF site; LEDGF site—*narrow*: docking was performed in a smaller region focused around the actual binding site of the ligands at the LEDGF site

Despite the difficulties, our virtual screening results were better than random, as shown by the enrichment factors and ROC curves and their area-under-curve (AUC) values (Table 3). It is interesting that the early enrichment obtained with the standard SIE function (EF of 1.25 at 10 % of ranked library, submission #146) was slightly improved with the application of the SIE + HB + FLAW scoring function

Fig. 5 ROC curves for virtual screening on HIV-integrase. The *curves* correspond to the models listed in Table 3

(1.79 at 10 % of ranked library, submission #147, see Fig. 5) overall the three sites. This contrasts with the slightly weaker performance of SIE + HB + FLAW versus SIE for pose prediction in the case of binders (Table 2), and hence indicates a role of the SIE + HB + FLAW function in filtering-out (via the E_{flaws} penalty) some of the false-positive non-binders. Encouraged by the docking results with the LEDGF box confined around the actual binding site, we retrospectively repeated virtual screening on this smaller box. While the performance was improved (Table 3; Fig. 5), this required prior knowledge of the system, which we purposely excluded from the blind evaluation of our methods, as it will not always be available in real-life applications of our tools.

In the case of alternate protonation and tautomeric sates of a ligand, Wilma–SIE selects the state with the lowest SIE value. Feasible alternate states were included prospectively for 13 compounds in the HIV-integrase virtual screening. Retrospectively, it turns out that all these compounds are non-binders, and they were correctly ranked low by Wilma–SIE. However, the RMS variation in SIE score between alternate forms is 0.96 kcal/mol, which is not negligible and underscores earlier reports of SIE sensitivity to protonation states [18]. There is a larger impact in the fragment site (1.28 kcal/mol) than in the LEDGF site (0.93 kcal/mol) or the Y3 site (0.51 kcal/mol). On the same subject, the selected alternate protonation of guest #10 of CB7 gave an improved SIE of −1.56 kcal/mol over the other protonation state. This translates into improving the correlation with experimental data, e.g., submission # 187

would have a decreased correlation coefficient of 0.71 from 0.74 (Table 1).

Conclusions

The SAMPL4 blind challenge provided a stringent test for the performance of the Wilma–SIE docking–scoring platform, which remains consistent with past experience on various systems. The strength of Wilma–SIE is in providing good correlations with binding affinities over dynamic ranges of 3 kcal/mol or wider. Using a common protein structure for all ligands can reduce the noise, while incorporating the more sophisticated solvation treatment of the FiSH model improves absolute predictions. Although the goal of consistently achieving sub-2 kcal/mol accuracy in relative binding free energies remains a challenge even when using the actual binding modes, the predictions correctly detect such narrow dynamic ranges. Estimation of the change in target's vibrational entropy may represent a way to improve absolute predictions. The present study further delineates the applicability domain of the Wilma–SIE platform for virtual screening. The formidable task of filtering out false positives may be improved by strengthening the penalty on non-complementary polar contacts. Wilma–SIE is not intended for detection of promiscuous weak binders with relatively high flexibility, although even in such difficult cases it can lead to better-than-random virtual screening results.

References

1. Reddy MR, Erion MD (2001) Free energy calculations in rational drug design. Springer, Berlin
2. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS (2011) Alchemical free energy methods for drug discovery: progress and challenges. Curr Opin Struct Biol 21:150–160
3. Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed 41:2644–2676
4. Gilson MK, Zhou HX (2007) Calculation of protein–ligand binding affinities. Annu Rev Biophys Biomol Struct 36:21–42
5. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL III (2004) Assessing scoring functions for protein–ligand interactions. J Med Chem 47:3032–3047
6. Wang R, Lu Y, Fang X, Wang S (2004) An extensive test of 14 scoring functions using the pdbbind refined set of 800 protein–ligand complexes. J Chem Inf Comput Sci 44:2114–2125
7. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006)

A critical assessment of docking programs and scoring functions. J Med Chem 49:5912–5931

8. Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR (2008) Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br J Pharmacol 153:S7–S26

9. Englebienne P, Moitessier N (2009) Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins? J Chem Inf Model 49:1568–1580

10. Purisima EO, Hogues H (2012) Protein–ligand binding free energies from exhaustive docking. J Phys Chem B 116:6872–6879

11. Chen W, Gilson MK, Webb SP, Potter MJ (2010) Modeling protein–ligand binding by mining minima. J Chem Theory Comput 6:3540–3557

12. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33:889–897

13. Gohlke H, Case DA (2004) Converging free energy estimates: MM-PB(GB)SA studies on the protein–protein complex Ras–Raf. J Comput Chem 25:238–250

14. Brown SP, Muchmore SW (2009) Large-scale application of high-throughput molecular mechanics with Poisson–Boltzmann surface area for routine physics-based scoring of protein–ligand complexes. J Med Chem 52:3159–3165

15. Naim M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, Siddiqi I, Drabik P, Sulea T, Bayly CI, Jakalian A, Purisima EO (2007) Solvated interaction energy (SIE) for scoring protein–ligand binding affinities. 1. Exploring the parameter space. J Chem Inf Model 47:122–133

16. Cui Q, Sulea T, Schrag JD, Munger C, Hung MN, Naim M, Cygler M, Purisima EO (2008) Molecular dynamics—solvated interaction energy studies of protein–protein interactions: the MP1-p14 scaffolding complex. J Mol Biol 379:787–802

17. Sulea T, Purisima EO (2012) The solvated interaction energy method for scoring binding affinities. Methods Mol Biol 819:295–303

18. Sulea T, Cui Q, Purisima EO (2011) Solvated interaction energy (SIE) for scoring protein–ligand binding affinities. 2. Benchmark in the CSAR-2010 scoring exercise. J Chem Inf Model 51:2066–2081

19. Skillman G. SAMPL1 at first glance. Cup IX meeting, Santa Fe, NM, 19 March 2008. http://eyesopen.com/2008_cup_presentations/CUP9_Skillman.pdf. Accessed 10 Jan 2014

20. Sulea T, Hogues H, Purisima EO (2012) Exhaustive search and solvated interaction energy (SIE) for virtual screening and affinity prediction. J Comput-Aided Mol Des 26:617–633

21. Dunbar JB, Smith RD, Yang C-Y, Ung PM-U, Lexa KW, Khazanov NA, Stuckey JA, Wang S, Carlson HA (2011) CSAR benchmark exercise of 2010: selection of the protein–ligand complexes. J Chem Inf Model 51:2036–2046

22. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197

23. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688

24. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general Amber force field. J Comput Chem 25:1157–1174

25. Purisima EO (1998) Fast summation boundary element method for calculating solvation free energies of macromolecules. J Comput Chem 19:1494–1504

26. Purisima EO, Nilar SH (1995) A simple yet accurate boundary element method for continuum dielectric calculations. J Comput Chem 16:681–689

27. Chan SL, Purisima EO (1998) Molecular surface generation using marching tetrahedra. J Comput Chem 19:1268–1277

28. Chan SL, Purisima EO (1998) A new tetrahedral tesselation scheme for isosurface generation. Comput Graph 22:83–90

29. Bhat S, Purisima EO (2006) Molecular surface generation using a variable-radius solvent probe. Proteins 62:244–261

30. 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 23:1623–1641

31. 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 21:132–146

32. 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 126:13156–13164

33. Chen W, Chang CE, Gilson MK (2004) Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. Biophys J 87:3035–3049

34. 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 6:1622–1637

35. Purisima EO, Corbeil CR, Sulea T (2010) Rapid prediction of solvation free energy. 3. Application to the SAMPL2 challenge. J Comput-Aided Mol Des 24:373–383

36. Ma D, Zavalij PY, Isaacs L (2010) Acyclic cucurbit[n]uril congeners are high affinity hosts. J Org Chem 75:4786–4795

37. McQuarrie DA (1976) Statistical mechanics. Harper & Row, New York

38. Gibb CL, Gibb BC (2004) Well-defined, organic nanoenvironments in water: the hydrophobic effect drives a capsular assembly. J Am Chem Soc 126:11408–11409

39. Sun H, Gibb CL, Gibb BC (2008) Calorimetric analysis of the 1:1 complexes formed between a water-soluble deep-cavity cavitand, and cyclic and acyclic carboxylic acids. Supramol Chem 20:141–147

40. 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 133:3570–3581

41. Moghaddam S, Inoue Y, Gilson MK (2009) Host–guest complexes with protein–ligand-like affinities: computational analysis and design. J Am Chem Soc 131:4012–4021

42. Gilson MK, Irikura KK (2010) Symmetry numbers for rigid, flexible, and fluxional molecules: theory and applications. J Phys Chem B 114:16304–16317

43. Gilson MK, Irikura KK (2013) Correction to "Symmetry numbers for rigid, flexible, and fluxional molecules: theory and applications". J Phys Chem B 117:3061

44. Peat TS, Rhodes DI, Vandegraaff N, Le G, Smith JA, Clark LJ, Jones ED, Coates JA, Thienthong N, Newman J, Dolezal O, Mulder R, Ryan JH, Savage GP, Francis CL, Deadman JJ (2012) Small molecule inhibitors of the LEDGF site of human immunodeficiency virus integrase identified by fragment screening and structure based design. PLoS ONE 7:e40147

45. Rhodes DI, Peat TS, Vandegraaff N, Jeevarajah D, Le G, Jones ED, Smith JA, Coates JA, Winfield LJ, Thienthong N, Newman J, Lucent D, Ryan JH, Savage GP, Francis CL, Deadman JJ (2011) Structural basis for a new mechanism of inhibition of HIV-1 integrase identified by fragment screening and structure-based design. Antivir Chem Chemother 21:155–168

46. Muddana HS, Fenley AT, Mobley DL, Gilson MK (2014) The SAMPL4 host–guest blind prediction challenge: an overview. J Comput-Aided Mol Des 28 (in press)