

Exhaustive search and solvated interaction energy (SIE) for virtual screening and affinity prediction

Traian Sulea · Hervé Hogues · Enrico O. Purisima

Received: 24 October 2011 / Accepted: 10 December 2011 / Published online: 25 December 2011
© Her Majesty the Queen in Rights of Canada as represented by the National Research Council of Canada 2011

Abstract We carried out a prospective evaluation of the utility of the SIE (solvation interaction energy) scoring function for virtual screening and binding affinity prediction. Since experimental structures of the complexes were not provided, this was an exercise in virtual docking as well. We used our exhaustive docking program, Wilma, to provide high-quality poses that were rescored using SIE to provide binding affinity predictions. We also tested the combination of SIE with our latest solvation model, first shell of hydration (FiSH), which captures some of the discrete properties of water within a continuum model. We achieved good enrichment in virtual screening of fragments against trypsin, with an area under the curve of about 0.7 for the receiver operating characteristic curve. Moreover, the early enrichment performance was quite good with 50% of true actives recovered with a 15% false positive rate in a prospective calculation and with a 3% false positive rate in a retrospective application of SIE with FiSH. Binding affinity predictions for both trypsin and host–guest complexes were generally within 2 kcal/mol of the experimental values. However, the rank ordering of affinities differing by 2 kcal/mol or less was not well predicted. On the other hand, it was encouraging that the incorporation of a more sophisticated solvation model into SIE resulted in better discrimination of true binders from binders. This suggests that the inclusion of proper Physics

in our models is a fruitful strategy for improving the reliability of our binding affinity predictions.

Keywords Virtual screening · Docking · SIE · Solvated interaction energy · Binding affinity prediction · FiSH

Introduction

The ability to accurately predict intermolecular associations is important for the understanding of the thermodynamic and structural aspects governing molecular recognition in biological systems. It is also critical to the success of important practical applications like the structure-based drug design. Hence, in the past two decades, the development of theoretical methods for predicting binding affinities has been fuelled by a perceived benefit to drug discovery. Binding affinity prediction methods span several levels of theory, with a corresponding trade-off between prediction accuracy and computational demand. On the one hand are the relatively slow but thermodynamically rigorous pathway approaches such as free energy perturbation (FEP) and thermodynamic integration (TI) [1, 2]. On the other hand is a large and ever-increasing number of faster approaches relying on binding affinity scoring functions that can be classified into three main categories: force-field-based, knowledge-based, and empirical [3–9].

An emergent group of end-point force-field based scoring functions that represent a reasonable compromise between time, computational resources, and accuracy combine molecular mechanics (MM) force-fields with a continuum treatment of solvation. A representative method in this group is MM-PB(GB)/SA [10–14], which combines MM-based terms with electrostatic solvation terms from generalized Born (GB) or Poisson–Boltzmann (PB) continuum models,

Electronic supplementary material The online version of this article (doi:10.1007/s10822-011-9529-7) contains supplementary material, which is available to authorized users.

T. Sulea · H. Hogues · E. O. Purisima (✉)
Biotechnology Research Institute, National Research Council
of Canada, 6100 Roylamount Avenue, Montreal,
QC H4P 2R2, Canada
e-mail: Enrico.Purisima@nrc.gc.ca

and surface area (SA)-based nonpolar solvation contribution. Solvated interaction energy (SIE) [15–17] is a similar end-point force-field-based scoring function that approximates the protein–ligand binding affinity by an interaction energy contribution and a desolvation free energy contribution, each of them further made up of electrostatic and nonpolar components. Electrostatic solvation effects are calculated with the boundary element solution to the Poisson equation, while non-polar solvation is based on molecular SA. Calibration of several physical parameters, including the dielectric constant, Born radii, surface tension coefficient, and enthalpy-entropy compensation scaling factor, was based on a diverse dataset of 99 protein–ligand complexes [15]. The SIE scoring function parametrized in this manner achieves a reasonable transferability across a wide variety of protein–ligand systems, consistently returning absolute binding affinities within the experimental range, as demonstrated by test cases published in the literature [18–31]. External testing of the standard SIE parametrization in the CSAR-2010 scoring challenge consisting in a curated dataset of 343 protein–ligand complexes diverse with respect to ligands and targets [32, 33], afforded binding affinity predictions with a mean-unsigned-error (MUE) of about 2 kcal/mol [34].

In this paper, we continue prospective testing of the SIE function. The first blind test was carried out in the first edition of SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) organized by OpenEye Scientific Software and showed a reasonable performance of SIE in binding affinity predictions for the SAMPL-1 set of kinase inhibitors for which available cognate crystal structures were provided [35]. However, the SAMPL-3 blind data sets propose significantly different challenges that test the limits of the applicability domain of the SIE function. First, the trypsin-binding fragments data set includes low-molecular-weight ligands that are typically of weak binding affinity (high- μM to mM) [36], a noisy range for most scoring functions. Secondly, the host–guest dataset challenges with systems in which the target is also of low-molecular-weight, but surprisingly, these systems have been considered notorious exceptions in the binding affinity landscape by having affinities unexpectedly high for their size [37, 38]. However, perhaps the most important test for the SIE function is the challenge in SAMPL-3 to work with non-experimental ligand poses for predicting binding affinities in both trypsin-fragment and host–guest systems. Clearly, the added challenge of scoring computationally derived binding modes is highly relevant for most of the real-life applications of SIE.

Therefore, a docking procedure was required in order to test SIE performance in SAMPL-3. We have recently developed Wilma (manuscript in preparation), an exhaustive docking program that has the required speed for large-

scale *in silico* docking-scoring (aka virtual screening) [39] 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 proven to be consistently close to the experimental pose for drug-like ligands. In SAMPL-3, the top-ranked Wilma pose(s) was (were) selected for post-scoring with SIE. In effect, here we test the performance of the Wilma-SIE docking-scoring platform for both virtual screening and binding affinity predictions.

Unquestionably, the success (or failure) of virtual screening (VS) relies mostly on the quality of the underlying docking and scoring function(s). The challenge in virtual screening is exacerbated by the fact that in order to be relevant in a drug discovery pipeline accurate docking-scoring has to be achieved under the constraint of fast computing. Because intermolecular binding is typically accompanied by the dehydration of the interacting surfaces and reorganization of the solvent water around the ensuing complex, a fast yet accurate solvation model is of paramount importance. This is afforded by the next-generation of solvation models that will retain the efficiency of the current continuum approximation but will be able to capture aspects of the physics of hydration that are dependent on the discrete properties of water. Such continuum models are the semi-explicit assembly (SEA) [40, 41], and first shell of hydration (FiSH) [42, 43]. Hence, we also used the SAMPL-3 data sets to test both prospectively and retrospectively the FiSH model, which we incorporated into the SIE function.

Methods

Wilma docking

Docking was carried out using an exhaustive docking software called Wilma (manuscript in preparation). 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 [44] (OpenEye Scientific Software, Santa Fe, NM) 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. The weighted 5-term scoring function used for docking was trained to recover the most native states using 343 protein–ligand complexes from the curated CSAR dataset [32]. The scoring function includes a van der Waals 6–12 Lennard–Jones potential, a Coulomb interaction term, an explicit H-bond term, which considers donor and acceptor orientations, and two surface and polar-surface complementarity

terms. Docking is done within a predefined rectangular volume 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 an internal energy cutoff of 20 kcal/mol and a minimal RMSD value that keeps the total number of conformations below 3,000 for the trypsin compounds or 10,000 for the larger host–guest ligands.

Solvated interaction energy (SIE) calculations

Scoring of binding affinities was carried out using the solvated interaction energy (SIE) end-point force-field based method [15–17, 34]. In SIE, the binding free energy in aqueous solution is approximated from the electrostatic and non-polar components of the interaction energy and the desolvation free energy (Eq. 1). The free state of the system is obtained from rigid separation of the interacting molecules from the bound state.

$$\text{SIE}(\rho, D_{\text{in}}, \alpha, \gamma, C) = \alpha [E_{\text{inter}}^{\text{Coul}}(D_{\text{in}}) + \Delta G_{\text{desolv}}^{\text{R}}(\rho, D_{\text{in}}) + E_{\text{inter}}^{\text{vdW}} + \gamma(\rho, D_{\text{in}}) \cdot \text{AMSA}(\rho)] + C \quad (1)$$

Intermolecular Coulomb and van der Waals interaction energies in the bound state, $E_{\text{inter}}^{\text{Coul}}$ and $E_{\text{inter}}^{\text{vdW}}$, were calculated with the biomacromolecular force field AMBER [45, 46], and its extension to small molecules, GAFF [47]. Partial atomic charges for protein atoms were taken from the AMBER force field, which are calculated with the two-stage RESP fitting method to the electrostatic potential at ab initio level [48, 49], whereas ligands were assigned AM1-BCC partial charges [50, 51]. For electrostatic desolvation, the change in the reaction field energy between the bound and free states, $\Delta G_{\text{desolv}}^{\text{R}}$, was calculated with a continuum model based on a boundary element solution to the Poisson equation using the BRI BEM program [52, 53]. The molecular surface required for boundary element electrostatic calculations was generated with a marching tetrahedra tessellation algorithm [54, 55], and a variable-radius solvent probe that adjusts with respect to the polarity of each atom being surfaced [56]. The generated molecular surface is also used to calculate the change in molecular surface area upon binding, MSA , leading to a nonpolar desolvation contribution upon multiplication with a surface tension coefficient, γ , which is based on a linear relationship between experimental hydration free energies of alkanes and their MSAs. ρ is a factor applied to derive atomic Born radii by linear scaling of AMBER van der Waals radii (R^*). D_{in} is the solute interior dielectric constant. 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 [57, 58].

Our main interest in SAMPL-3 was to test prospectively the default values of $\rho = 1.1$, $D_{\text{in}} = 2.25$, $\gamma = 12.894$ cal/(mol Å), $\alpha = 0.104758$, and $C = -2.89$ kcal/mol, which represent the standard SIE parameters originally obtained by calibration against a protein–ligand training dataset of 99 complexes refined by restrained energy minimization [15]. We also explored prospectively rescaled SIE functions where the α and C parameters were retrained on published data for SAMPL-3 systems. For trypsin affinity prediction, we rescaled the SIE function using a subset of 16 trypsin-ligand complexes from the original SIE training data set [15]. This resulted in values of $\alpha = 0.1609$ and $C = 2.16$ kcal/mol, specifically tuned for trypsin. For host–guest affinity predictions, SIE rescaling was based on free energy data available for 26 guests binding to host 1 and 7 guests binding to host 2 [59]. We note that SIE rescaling affects absolute predictions (e.g., MUEs) but not the level of correlation between experimental and SIE-predicted binding affinities. We shall refer to the rescaled SIE function as rSIE.

Single-conformation-based SIE calculations were performed on complexes refined by constrained energy minimization [15]. In the case of trypsin-ligand complexes, we applied our current refinement protocol for protein–ligand complexes [34, 35], which includes energy minimization of the ligand and only the protein residues within 4 Å from the ligand, and applying harmonic restraints with force constants of 3 kcal/(mol Å²) and 20 kcal/(mol Å²) for the ligand and protein, respectively, heavy atoms in this region. For the host–guest systems, the harmonic restraints were applied on all heavy atoms in the system, 3 kcal/(mol Å²) for the guest and 20 kcal/(mol Å²) for the host. Energy minimization was carried out down to a gradient of 0.01 kcal/(mol Å), with AMBER/GAFF force-field parameters and two-stage-RESP/AM1-BCC partial charges (as employed in SIE calculations), and a distance-dependent dielectric constant ($4r$) to crudely mimic solvent screening.

FiSH solvation model

The FiSH solvation model was designed to capture some of the discrete nature of hydration within a completely continuum framework [42]. By using Born radii that depend on the induced surface charge density it reproduces the charge asymmetry of hydration observed in discrete water simulations [60, 61]. 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. Furthermore, the non-electrostatic component of solvation is split into a cavity term and a solute–solvent van der Waals term. The non-bulk

nature of the first hydration shell is represented by a two-region continuum van der Waals model. Water in the first hydration shell is modeled as a uniform distribution along the solvent-accessible surface (SAS) constructed using atom-dependent probe radii. The van der Waals interaction of the solute with the first hydration shell is calculated by integrating the Lennard–Jones potential along the SAS [42] using AMBER [45, 46] or GAFF [47] parameters for the solute and TIP3P [62] parameters for water. The van der Waals contribution of the second solvation shell outwards is obtained by integrating the contribution of a uniform bulk solvent from the SAS + 2.8 Å out to infinity using standard continuum van der Waals methods [42, 63].

Structural preparation

Trypsin data set

Three high-resolution crystal structures of bovine trypsin were prepared for virtual screening, PDB entries 1HJ9, 1SOR and 3MI4 refined at resolutions of 0.95, 1.02, and 0.8 Å, respectively. A superposition of these structures reveals only minor structural deviations around the active site; however, the Gln192 side chain located at the opening of the S1 pocket adopts different rotameric states. Structural preparation was done in SYBYL 8.1.1 (Tripos, Inc., St. Louis, MO). Bound ligands and buffer ions were removed. Hydrogen atoms were added, with the ionizable groups protonated at neutral pH. Tautomeric and protonation states of His residues were manually assigned after visual inspection in order to maximize the H-bonding network. A Ca²⁺ ion distant from the catalytic site was retained. With respect to the treatment of crystallographic water molecules, we prepared two versions for each structure. In one version, all explicit solvent molecules were removed. In another version we retained 22 water molecules conserved among the three crystal structures used, 14 of which are buried in the protein core, 3 are proximal to Ca²⁺, 4 are buried in the back of the Asp189 side chain at the bottom of the S1 pocket, and 1 bridges the main-chain atoms of residues Ser217, Gln221, and Lys224 in the wall of the S1 pocket. Polar hydrogen atoms were manually oriented to maximize H-bonding. All prepared trypsin structures were then subjected to energy minimization with the AMBER force-field, in which all hydrogen atoms were allowed to move with heavy atoms fixed at their crystallographic positions.

In order to prepare the fragments database for virtual screening, we first assigned the protonation states of the 544 ligands in the database at neutral pH using FILTER (OpenEye Inc., Santa Fe, NM). Manual changes were made in the protonation states produced by FILTER for 11 ligands. These included migration of the proton from the more buried amine to the more exposed amine in the piperazine moieties of

ligands ID 113, 114, 215, 216, 217, 245, 304, 330, 356, as well as protonation at the exposed N atom of the hydrazine moiety in ligand ID 178, and protonation of the tertiary aliphatic amine in the ligand ID 488. Partial charges were calculated with the AM1-BCC method [50, 51], as implemented in MOLCHARGE (OpenEye, Inc.), using as input the lowest-energy conformation generated by OMEGA (OpenEye, Inc.). The same preparation procedure of the target and ligands was used for trypsin binding affinity prediction.

Host–guest data set

Host-1, an acyclic cucurbit[*n*]uril (CB[*n*]) congener containing 4 carboxylate side chains, was prepared in two conformations starting from the high and low occupancy states observed crystallographically in the bound state with a linear aliphatic tetramine guest [59]. For the prospective study, two protonation states were considered for host-1, with the carboxylate groups ionized and neutral. Each host 1 structural variant was energy minimized with the GAFF force field [47], AM1-BCC partial charges, a 4 r distance-dependent dielectric and harmonic force restraints on the heavy atoms of 20 kcal/(mol Å). Provided structures for host-2 and host-3, the neutral cyclic CB (7) and CB (8) hosts respectively, were energy minimized with the same settings as host-1, except that no restraints were imposed. The structures of the 7 guests binding to host-1, and the 2 guests binding to each of host-2 and host-3, were protonated at neutral pH and partial charges calculated as described earlier for trypsin ligands. A training set of guests with measured binding affinities, comprising 26 guests binding to host-1 and 7 guests binding to host-2 [59], was also prepared in the same manner.

Results

Trypsin virtual screening

We submitted two prospective predictions for trypsin virtual screening. The two submissions differed in the way the predicted pose was selected for scoring. In one set, the pose with the best Wilma docking score for each ligand was used and subjected to restrained energy minimization (see “Methods” section) and rescored using the SIE energy function with default parameters. We will refer to that pose as the Top-Wilma pose. In the second approach, the poses with the top 100 Wilma docking scores for each ligand were clustered and representatives of each cluster were subjected to restrained energy minimization followed by SIE rescoring. The best SIE score among them was selected as the virtual screening score for the ligand. We will refer to the associated pose as the Top-SIE pose.

We used three crystal structures of trypsin (pdb codes 1HJ9, 1SOR and 3MI4) as targets for docking. The submitted predictions were based on trypsin structures with the crystallographic water molecules removed. We also carried out the calculations with selected conserved water molecules retained. However, the results were highly correlated with those for the dry trypsin structures and we opted to base all our submissions on the dry trypsin targets. For each of these targets, a rectangular box ($23.7 \text{ \AA} \times 18.0 \text{ \AA} \times 29.0 \text{ \AA}$) enclosing the substrate-binding groove of trypsin (Fig. 1) was defined as the relevant region for exhaustive virtual docking using Wilma. In general, the top-scoring poses were docked at the S1 specificity pocket of trypsin. Each ligand was assigned the best score obtained across the three trypsin structures.

The SAMPL-3 virtual screening set consisted of 544 compounds of which 20 were true binders. Figure 2 shows the receiver operating characteristic (ROC) curve for these two sets of predictions. The performance of the two methods is very similar with the Top-SIE poses giving a somewhat better area under the curve (AUC). The bootstrapped AUCs are 0.70 and 0.68 for the Top-SIE and Top-Wilma poses, respectively. (The perfect AUC would have a value of 1, indicating all true binders are ranked at the top of the list; a random ranking would give an AUC of 0.5.) AUCs are sensitive to false negatives that are detected only late in the screening process. We have three true binders that are ranked near the bottom of the list. These three false negatives alone result in about a 10% reduction in the AUC. It should be noted that the early enrichment performance is quite good, with 50% of the true binders obtained with a 15.6% false positive rate for the Top-SIE set.

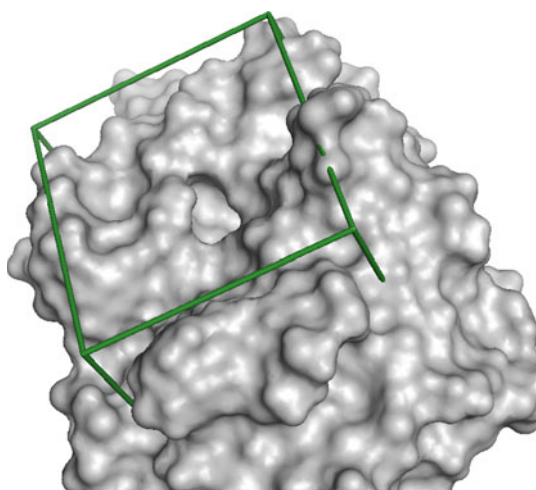


Fig. 1 Box defining the region used for virtual docking. Trypsin is represented as a molecular surface. The box includes the S1 pocket as well as a large part of the binding groove around it

Trypsin affinity prediction

We submitted several prospective models of trypsin affinity predictions. These are summarized in Table 1 along with the statistical measures of their performance. Aside from testing the effect of which docked pose (Top Wilma or Top SIE) to score for affinity we also tested three scoring functions. These were (a) the SIE function with default parameters, (b) the rSIE (rescaled SIE) function with parameters $\alpha = 0.1609$ and $C = 2.16$ in Eq. 1 that were optimized for trypsin and (c) SIE + FiSH, an SIE function with the solvation model replaced by the FiSH solvation model. The calculations carried out for affinity prediction were exactly the same as those used for the virtual screening exercise except for the additional scoring functions tested. As in the virtual screening case, there was not much difference between using Top-Wilma poses versus Top-SIE poses, although the latter performed slightly better. For the discussion that follows, we will focus on the results using the Top-SIE poses. Figure 3 shows scatter plots comparing the predicted and experimental binding affinities for each of the three scoring functions using the Top-SIE poses. The 34-compound set was composed of 17 binders and 17 non-binders. For the purpose of analyzing the results, the non-binders have been arbitrarily given an “experimental” value of -4.09 kcal/mol. Compared to the default SIE the use of rSIE improved the agreement of the predicted and experimental affinities but does not in any way alter the relative ranking of affinities. With rSIE, most of the predicted affinities for true binders are within 2 kcal/mol of the experimental values. The mean unsigned error (MUE) and median unsigned error (MdUE) are 0.64 and 0.30 kcal/mol, respectively. However, the correlation coefficients are rather poor, $r^2 = 0.00$ and Kendall $\tau = 0.12$. For the purpose of the statistical analysis, binding affinities of non-binders that are predicted to be more positive than -4.09 kcal/mol have been capped at that value to equal the “experimental” value assigned to non-binders. Restricted to the true binders, the MUE and MdUE are 1.10 and 0.76 kcal/mol, respectively.

The FiSH solvation model [42, 43] is more sophisticated than the default solvation model of SIE. Instead of a single surface-area-based term for the non-electrostatic component of solvation, it includes additional terms for a continuum van der Waals representation of solute–solvent interactions. The modified SIE + FiSH scoring function then has the form

$$\text{SIE} + \text{FiSH} = \alpha \cdot (E_{\text{coul}} + E_{\text{RF}} + E_{\text{vdW}} + E_{\text{cvdW}} + E_{\text{cav}}) + C \quad (2)$$

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

Fig. 2 ROC curves for virtual screening on trypsin. **a** Using Top-SIE pose. **b** Using Top-Wilma pose

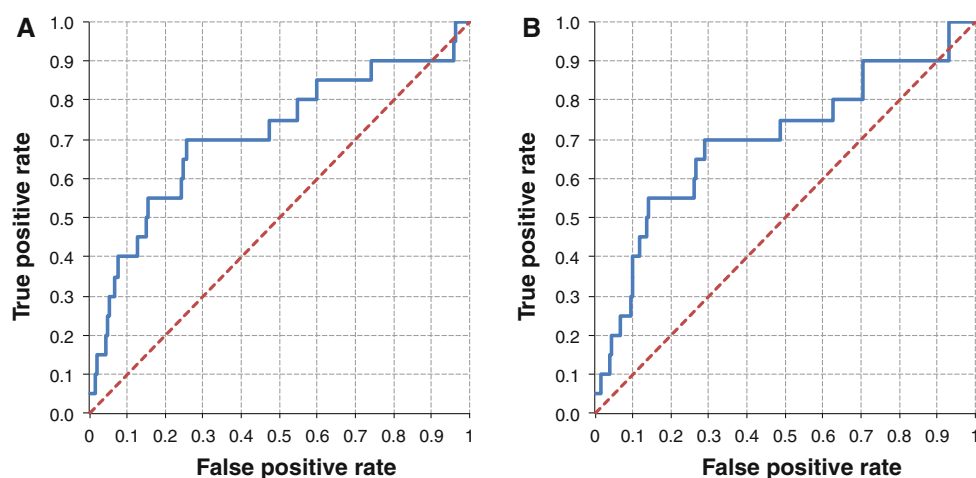


Table 1 Statistical performance of trypsin fragment affinity prediction models

Model			MSE	MUE	MdUE	RMSE	$K\tau$	r^2	Slope	Intercept
Official ID#	Pose	Scoring								
86	Top-SIE	SIE + FiSH	-0.72	0.82	0.25	1.71	0.14	0.03	-0.02	-4.41
			<i>0.26</i>	<i>0.26</i>	<i>0.14</i>	<i>0.50</i>	<i>0.17</i>	<i>0.05</i>	<i>0.14</i>	<i>0.55</i>
87	Top-SIE	SIE	2.11	2.14	2.27	2.32	0.30	0.12	0.20	-3.19
			<i>0.17</i>	<i>0.16</i>	<i>0.24</i>	<i>0.14</i>	<i>0.11</i>	<i>0.07</i>	<i>0.08</i>	<i>0.52</i>
88	Top-SIE	rSIE	-0.32	0.64	0.30	1.14	0.12	0.00	0.03	-4.41
			<i>0.19</i>	<i>0.16</i>	<i>0.13</i>	<i>0.28</i>	<i>0.15</i>	<i>0.04</i>	<i>0.14</i>	<i>0.59</i>
89	Top-Wilma	SIE	2.09	2.11	2.23	2.32	0.22	0.08	0.15	-3.48
			<i>0.17</i>	<i>0.17</i>	<i>0.25</i>	<i>0.15</i>	<i>0.12</i>	<i>0.06</i>	<i>0.08</i>	<i>0.49</i>
90	Top-Wilma	SIE + FiSH	-0.92	1.02	0.36	1.85	-0.02	0.01	-0.02	-4.59
			<i>0.28</i>	<i>0.27</i>	<i>0.22</i>	<i>0.46</i>	<i>0.18</i>	<i>0.05</i>	<i>0.07</i>	<i>0.23</i>
91	Top-Wilma	rSIE	-0.38	0.74	0.31	1.21	0.08	0.02	0.04	-4.49
			<i>0.20</i>	<i>0.17</i>	<i>0.17</i>	<i>0.27</i>	<i>0.14</i>	<i>0.03</i>	<i>0.09</i>	<i>0.38</i>
Null ^a			NA	0.45	0.42	0.59	NA	NA	NA	NA

Median and standard error (italicized) values from bootstrapping using 5,000 replicates with replacement

MSE mean signed error, MUE mean unsigned error, MdUE median unsigned error, RMSE root-mean-square error, $K\tau$ Kendall tau, r^2 Pearson correlation coefficient

^a Null hypothesis, i.e., predicted affinities set to the experimental mean

As with the rSIE case, most of the predicted affinities for true binders are within 2 kcal/mol of the experimental values. The MUE and MdUE are 0.81 and 0.25 kcal/mol, respectively. However, the correlation coefficients are rather poor, $r^2 = 0.00$ and Kendall $\tau = 0.14$. Restricted to the true binders, the MUE and MdUE are 1.57 and 0.77 kcal/mol, respectively.

The overall performance of SIE + FiSH seems to be similar to that of re-scaled SIE. However, compared to rescaled SIE, SIE + FiSH appears to discriminate true binders from non-binders (Fig. 3a, c) better. We see that for rescaled SIE, the range and spread of values for the non-binders is similar to that of the true binders. With SIE + FiSH, the true binders tend to be more negative than

the non-binders. Given this observation, we applied the SIE + FiSH scoring function retrospectively to the VS data set. The result is a dramatic improvement in the early enrichment performance (Fig. 4). For SIE, 50% of the true positives were obtained with a 15% false positive rate (Fig. 2). With SIE + FiSH, 50% of true positives were obtained with a 3% false positive rate. However, the AUC is only slightly increased due to the large penalty for the three false negatives that are ranked close to the bottom.

Host-guest affinity prediction

We submitted several prospective models of host-guest affinity predictions. These are listed in Table 2 along with

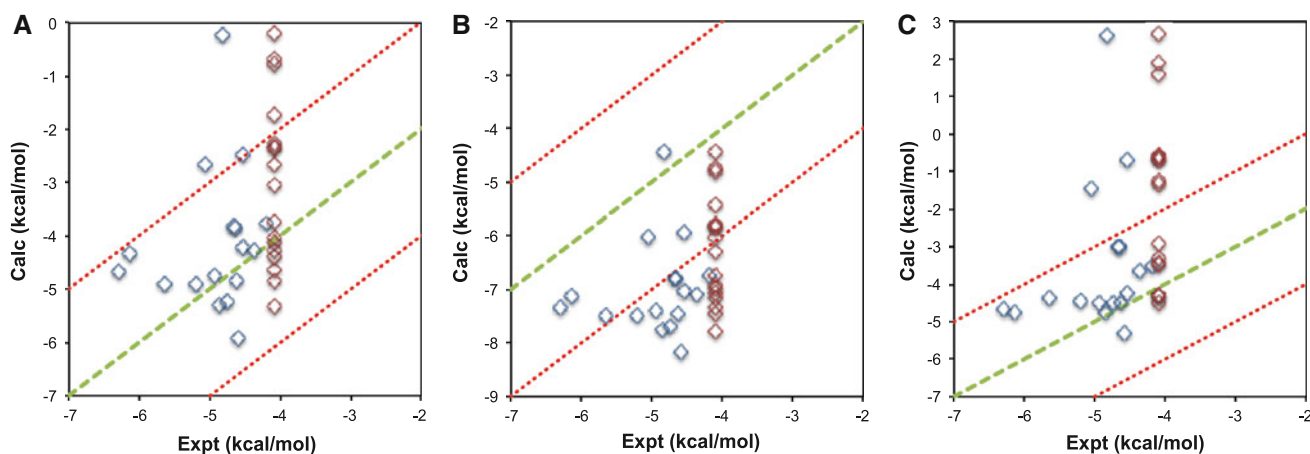


Fig. 3 Scatter plot of predicted versus experimental binding affinities of trypsin ligands. **a** Rescaled SIE parameters. **b** Default SIE parameters. **c** SIE + FiSH. The “experimental” value for non-binders (red diamonds) has been arbitrarily set to -4.09 kcal/mol.

The dashed line represents perfect correlation. Points between the dotted lines have predicted affinities within ± 2 kcal/mol from the experimental value

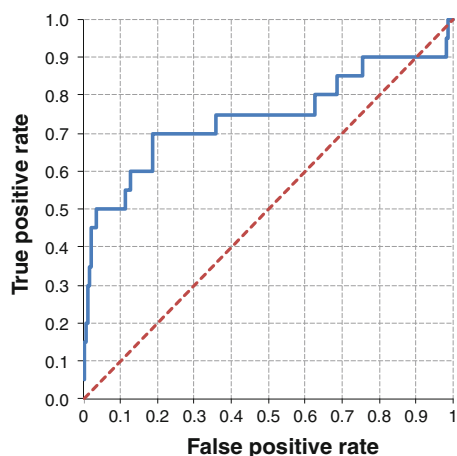


Fig. 4 ROC curve for virtual screening on trypsin using SIE + FiSH

their statistical performance on the combined data set of 11 host–guest complexes (7 guests for host-1, and 2 guests for each of host-2 and host-3). The affinities predicted for these complexes with the models listed in Table 2 are provided in Table S1. Host-1 is an acyclic cucurbituril (CB) analog that is ionizable due to its four carboxylate side chains, whereas host-2 and host-3 are cyclic CB (7) and CB (8) analogs which are neutral. All these hosts have a circular geometry with a central hole where certain guests are recognized with surprisingly high affinity given the relatively small size of these systems [59]. 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 fully or partially through the central hole region of the hosts (Fig. 5), irrespective of the structural setup (neutral/ionized, high/

low occupancy conformation) or pose-scoring function (Wilma, SIE, SIE + FiSH). We also docked a set of 26 guests to host-1 and 7 guests to host-2 with published binding affinities [59], with the intention of rescaling the SIE function specifically for host–guest systems. The training set of guests also docked in the central region of the hosts, and the top-scored pose was found to be similar to the binding modes previously determined experimentally for two of these guests (Figure S1) [59].

All models submitted to SAMPL-3 host–guest affinity prediction challenge are based on the high-occupancy conformation of host-1, because prospectively very similar results were obtained when the low-occupancy conformation was used ($r^2 > 0.98$, mean-unsigned-deviation < 0.2 kcal/mol). An exception is guest-3, which in some models was predicted to bind more weakly to the low-occupancy conformation than to the high-occupancy one. Guest-3 is a branched and relatively larger guest in this series and Wilma mainly employs a rigid docking algorithm. Also, on a previously published data of 26 guests binding to host-1 [59], we obtained better correlations with experiment using the high-occupancy conformation (data not shown). All these prompted us to discard data generated on the low-occupancy conformation of host-1.

The prediction model #94 is based on the Top-SIE poses. Predictions for host-1 were generated with this host in the fully ionized form (net charge of $-4e$). This model returned a reasonable correlation with experiment ($K\tau$ of 0.49; r^2 of 0.51) and also a good prediction of absolute binding affinities as shown by the MUE of 1.21 kcal/mol and RMSE of 1.54 kcal/mol, which are better than the null model for this data set (MUE = 1.44 kcal/mol and RMSE = 1.79 kcal/mol). We note, however, the relatively

Table 2 Statistical performance of host–guest binding affinity prediction models

Model				MSE	MUE	MdUE	RMSE	K τ	R ²	Slope	Intercept
Official ID#	Pose	Scoring	Host-1 state								
94	Top SIE	SIE ^c	Ionized	0.88	1.21	1.02	1.51	0.49	0.50	0.42	−3.40
				<i>0.38</i>	<i>0.29</i>	<i>0.50</i>	<i>0.29</i>	<i>0.24</i>	<i>0.22</i>	<i>0.15</i>	<i>1.15</i>
96	Top W	SIE	Ionized	1.16	1.36	1.25	1.67	0.51	0.52	0.45	−2.85
				<i>0.37</i>	<i>0.31</i>	<i>0.51</i>	<i>0.29</i>	<i>0.26</i>	<i>0.23</i>	<i>0.15</i>	<i>1.11</i>
101	Top SIE	SIE	Neutral	0.58	1.17	0.99	1.46	0.41	0.50	0.29	−4.69
				<i>0.42</i>	<i>0.28</i>	<i>0.33</i>	<i>0.34</i>	<i>0.21</i>	<i>0.22</i>	<i>0.10</i>	<i>0.84</i>
99	Top W	rSIE ^d	Neutral	−2.11	2.11	1.81	2.48	0.45	0.49	0.70	−4.37
				<i>0.43</i>	<i>0.43</i>	<i>0.46</i>	<i>0.47</i>	<i>0.22</i>	<i>0.23</i>	<i>0.25</i>	<i>2.05</i>
Retro ^a	Top SIE	rSIE ^e	Ionized	0.19	1.27	1.12	1.56	0.49	0.49	0.83	−1.09
				<i>0.48</i>	<i>0.29</i>	<i>0.45</i>	<i>0.29</i>	<i>0.25</i>	<i>0.22</i>	<i>0.30</i>	<i>2.35</i>
98	Top SIE	SIE + FiSH ^f	Ionized	5.58	5.58	5.42	5.93	0.34	0.30	0.62	2.77
				<i>0.64</i>	<i>0.64</i>	<i>1.05</i>	<i>0.67</i>	<i>0.27</i>	<i>0.21</i>	<i>0.40</i>	<i>3.10</i>
100	Top SIE	rSIE + FiSH ^g	Ionized	−0.99	1.59	1.65	1.82	0.34	0.31	0.33	−5.92
				<i>0.47</i>	<i>0.28</i>	<i>0.42</i>	<i>0.26</i>	<i>0.27</i>	<i>0.22</i>	<i>0.21</i>	<i>1.63</i>
Null ^b				NA	1.44	1.28	1.79	NA	NA	NA	NA

Data shown for the combined set of 11 host–guest complexes (7 guests for host 1 and 2 guests for each of host 2 and host 3). Median and standard error (italicized) values from bootstrapping using 5,000 replicates with replacement

MSE mean-signed-error, *MUE* mean-unsigned-error, *MdUE* median-unsigned-error, *K τ* Kendall's tau rank-order correlation coefficient, *r*² Pearson squared correlation coefficient, *Top SIE* selection of top-ranked pose based on standard SIE scoring function, *Top W* selection of top-ranked pose based on Wilma scoring function, *SIE* standard SIE function, *rSIE* rescaled SIE function based on an external training data set, *SIE + FiSH* standard SIE + FiSH function, *rSIE + FiSH* rescaled SIE + FiSH function based on an external training data set

^a Retrospective model. SIE rescaled on the same external training data set as used prospectively for the model #99 and with host 1 in ionized state

^b Null model given by the arithmetic average of experimental data

^c $\alpha = 0.1048$; $C = -2.89$ (see Eq. 1)

^d $\alpha = 0.2568$; $C = 0$ (see Eq. 1)

^e $\alpha = 0.2097$; $C = 0$ (see Eq. 1)

^f $\alpha = 0.1232$; $C = 1.46$ (see Eq. 2)

^g $\alpha = 0.06497$; $C = -6.64$ (see Eq. 2)

small correlation slope (0.441) and significant correlation intercept (−3.22), which underscore the narrower range of predicted absolute binding affinities than the experimental range, also apparent in the scatter plot in Fig. 6. For most complexes, the standard SIE function slightly underestimated absolute binding affinities, leading to a positive MSE value of 0.88 kcal/mol.

The prediction model #96 is similar to model #94, with the only difference being that it uses the Top-Wilma poses. Note that final scoring in both models is based on the standard SIE function. Significant prediction differences were observed for only 2 complexes of host-1, a 1.5 kcal/mol more positive SIE value (weakened predicted binding) for guest-3, and a 0.5 kcal/mol weakening in the case of guest-7. The difference in the selected poses for these two complexes by the two methods is shown in Fig. 7. In both cases, the SIE values based on Wilma pose selection (model #96) are farther from experimental values than

predictions based on SIE pose selection (model #94), which is reflected in the slightly larger MUE and RMSE values. However, correlation parameters ($K\tau$, r^2 , slope, intercept) for model #96 improve marginally relative to model #94 (Table 2).

We also tested prospectively the protocol from model #94 against a protonated (neutral) version of host-1 (prospective prediction #101, Table 2). Somewhat to our surprise, this model was our best submission in terms of absolute predictions, with the MUE as low as 1.16 kcal/mol and RMSE of 1.49 kcal/mol. Correlation indices deteriorate slightly relative to model #94, but still provide a similar r^2 of 0.50. For most guests, SIE values are slightly more negative (stronger predicted binding) for the neutral host-1 (model #101) than for the ionized host-1 (model #94), by as much as 1.25 kcal/mol in the case of guest-7, with the only exception being guest-6 having weakened predicted binding by 0.45 kcal/mol. To provide a

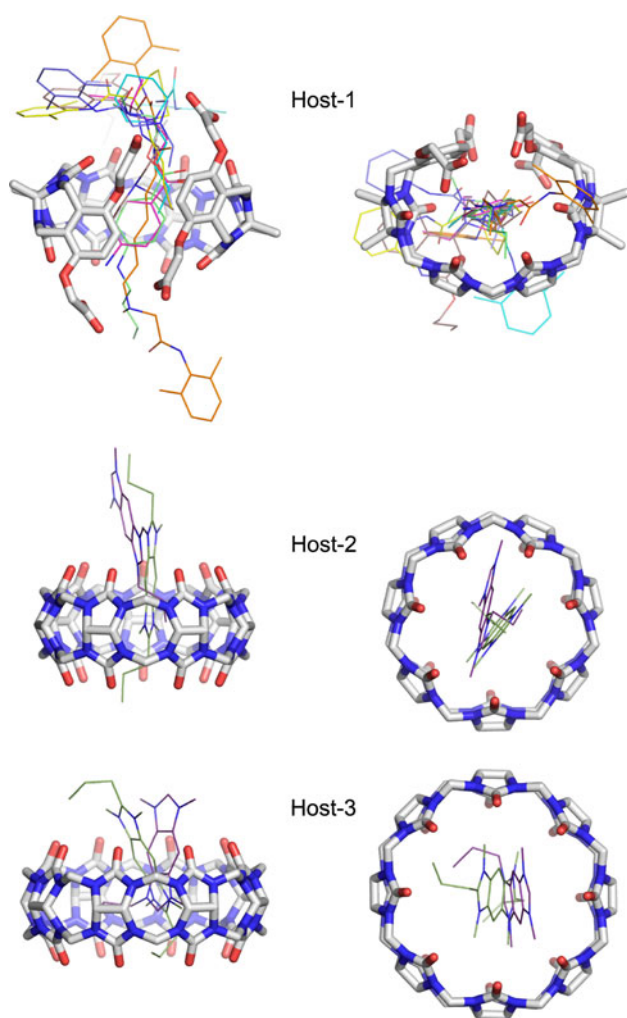


Fig. 5 Global views of Wilma docking poses for the host–guest systems. The examples shown are from the SIE-based selection of the top-ranked pose and the ionized host-1. All guests are shown in a given host with *different colors* of the C atoms. Two orthogonal views are provided

qualitative view of differences in the interactions, in Fig. 8 we display the top-ranked poses of several guests with host-1 in the ionized and neutral states. We note that the correlation slope for this model has decreased to a low value of 0.290. As seen in the scatter plot in Fig. 6, the predicted binding affinities span 2.5 kcal/mol whereas the experimental values range over 6.5 kcal/mol.

One way to modulate the correlation slope is to rescale the SIE function in terms of the enthalpy-entropy compensation factor α in Eq. 1 specifically for the system being investigated. This is justified since it has been previously shown that the CB (7) 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 (β CD) host [37, 38, 58]. This points

towards a larger value for the α scaling factor in the SIE formulation. Hence, we explored this possibility prospectively by deriving a rescaled SIE function based on a previously published data for guests binding to host-1 (26 complexes) and host-2 (7 complexes) [59]. Amongst the many system and method variants tested in the prospective analysis (neutral/ionized host, Wilma/SIE pose selection, high/low occupancy conformations) the best fit was obtained for the Wilma-based selection of the pose and the neutral host-1. This training model achieved an MUE of 1.56 kcal/mol over all 26 guests for host-1 and 7 guests of host-2, and led to an α scaling factor of 0.2568 (the constant C was forced to zero), hence larger scaling than that for the standard SIE function (0.1048), in agreement with previous observations [37, 38]. Application of the rescaled SIE function to the SAMPL-3 host–guest data set led to the submitted prediction model #99, with an increased correlation slope (0.705) and similar correlation with experiment relative to the other prospective models based on standard SIE function. Two aspects in terms of absolute affinity prediction are noteworthy: rescaling led to overshooting of predictions from underestimating to overestimating actual affinities (negative MSE, see also Fig. 6), and increase of MUE to above 2 kcal/mol (Table 2).

Therefore, we retrospectively reanalyzed our un-submitted prospective models of rescaled SIE function on the training set of 33 complexes of host-1 and host-2 [59]. A particularly interesting model turns out to be the one employing Top-SIE pose selection and ionized host-1. This model was not submitted prospectively because it performed poorer than model #99 in the training stage, with a training MUE of 2.13 kcal/mol (versus 1.56 kcal/mol for model #99). Its α scaling factor of 0.2097 is larger than that in the standard SIE function used in model #94 that underestimated the actual data, but smaller than in the SIE function rescaled on the training dataset with neutral host-1 used in model #99 that overestimated actual affinities. As seen in Table 2 and Fig. 6, this retrospective model (rescaled SIE, ionized host-1) has a much improved correlation slope (0.880) relative to model #94 (standard SIE, ionized host-1) and an improved MUE (1.27 kcal/mol) relative to model #99 (rescaled SIE, neutral host-1). The correlation coefficient is similar to the other models, but the correlation intercept and the MSE are much improved being close to zero.

We also tested SIE + FiSH. The SAMPL-3 prospective prediction model #98, which is based on SIE + FiSH final scoring and ionized host-1, did not perform well, showing much deteriorated correlation and absolute affinity prediction relative to the standard SIE function (Table 2; Fig. 6). The severe underestimation of binding affinities by this model (MSE = MUE of 5.59 kcal/mol) is corrected in the prospective model #100, which uses a rescaled SIE/

Fig. 6 Scatter plots of calculated versus actual binding affinities for the host–guest systems and various prediction models. Please refer to Table 2 for a description of the models. Host-1 data are shown with *blue circles*, host-2 data with *red triangles* and host-3 data with *yellow squares*. The *diagonal line* indicates a perfect correlation. Select outliers discussed in the text are labeled

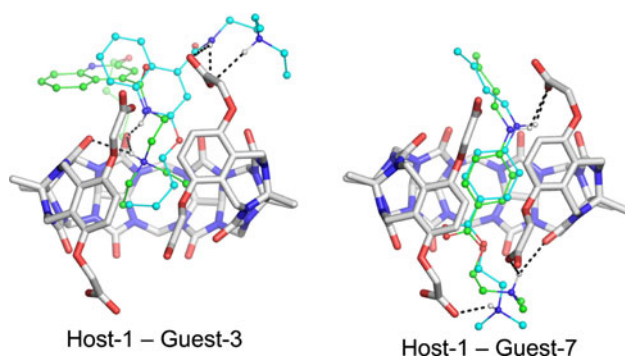
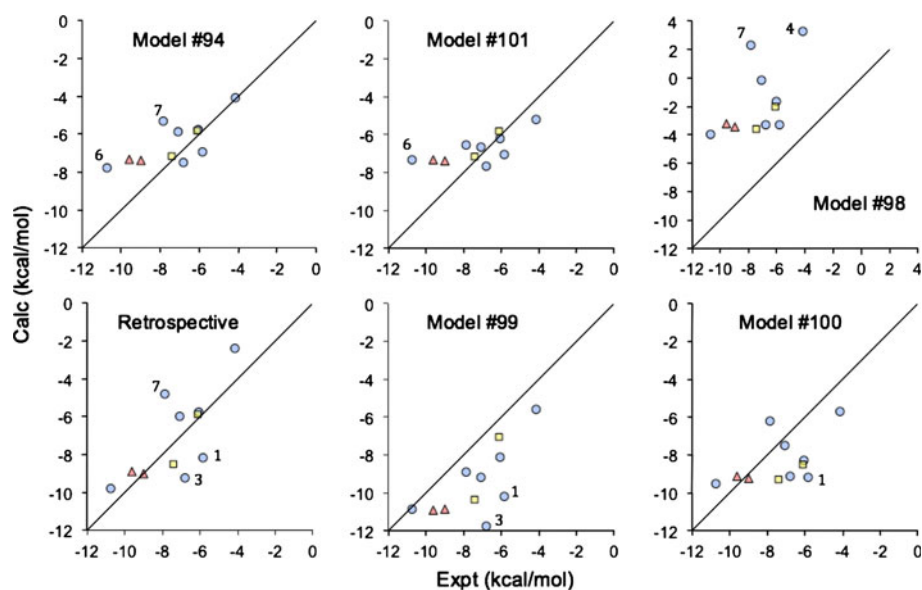


Fig. 7 Differences in the Top-Wilma and Top-SIE pose selection. The shown complexes are the only ones that have significantly different predicted binding affinities and different selected poses by the two scoring functions; see Table S1 and the text for the difference in predicted binding affinities. The C atoms of the guests (*ball-and-stick* models) are *colored in green and cyan* for the SIE-selected and Wilma-selected poses, respectively. H-bonds are indicated as *black dashed line*. Only polar H atoms are shown for clarity

FiSH model derived on the training set of 33 complexes of host-1 and host-2 [59], which cannot however improve the correlation with experiment. Obviously, more work is needed for a consistent incorporation the FiSH solvation model into the SIE function, at least for the host–guest systems examined here.

Discussion

Outliers of trypsin virtual screening and affinity prediction

Figure 9 shows chemical structures of three serious outliers in our affinity prediction results. These same outliers also

adversely affected the AUC in the virtual screening results. Figure 10 also shows the predicted binding mode of one of the outliers, frag.aff.15. In this pose, the imidazo nitrogen on the ligand points away from Asp189 and towards the backbone NH of Ser214. However, at a distance of 3.4 Å from the amide hydrogen, it is too far to form a good hydrogen bond. This pose suggests that if the imidazo nitrogen were protonated the ligand could flip and form a stabilizing ion pair with Asp189 of trypsin. By analogy with other imidazo compounds, it is plausible that the imidazo nitrogen in frag.aff.15 is protonated near neutral pH. Figure 10 (right panel) shows the predicted pose for the protonated version of frag.aff.15. After protonation, the predicted binding affinity using the rSIE scoring function goes from -4.45 to -6.44 kcal/mol with VS ranking going from 522 to 178. For the SIE + FiSH scoring function, the predicted binding affinity goes from 2.63 to -2.78 kcal/mol with VS ranking going from 534 to 124. The second outlier, frag.aff.16, also has an imidazole nitrogen that was not protonated. For rSIE, protonation changes the predicted binding affinity from -2.64 to -3.77 kcal/mol and raises the VS rank from 262 to 147. For SIE + FiSH, protonation changes the predicted binding affinity from -1.44 to -3.37 kcal/mol and raises the VS rank from 203 to 90. The docked conformation of the third outlier, frag.aff.27 (Fig. 9), had the aniline and pyrrole rings nearly perpendicular to each other. By docking a conformation in which the two rings are nearly planar, the predicted binding affinity and rank improved only marginally. The rSIE affinity went from -5.02 to -5.96 kcal/mol with rank rising from 302 to 285. For SIE + FiSH, the predicted affinity went from -0.69 to -0.79 kcal/mol with rank rising from 343 to 330. The corrected outliers improved the AUC for SIE + FiSH from 0.73 to 0.78 (Fig. 11).

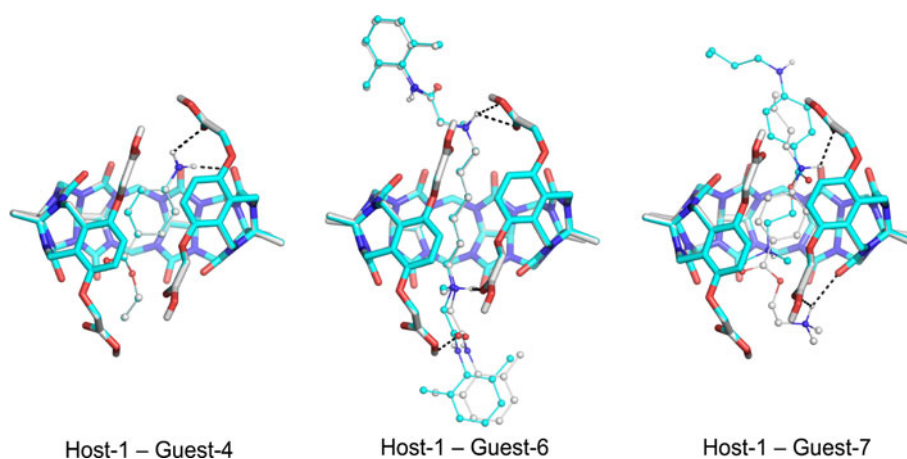


Fig. 8 Differences in the top-pose interactions with the ionized vs neutral states of host-1. See Table S1 and the text for the difference in predicted binding affinities for each shown example. Top-pose selection is based on the SIE function on both cases. The C atoms of the guests (ball-and-stick models) are colored in grey and cyan for

the poses bound to the ionized and neutral host-1 states, respectively. H-bonds are indicated as *black dashed line*. Only polar H atoms are shown for clarity. Note that the ionized and neutral host structures overlay almost perfectly

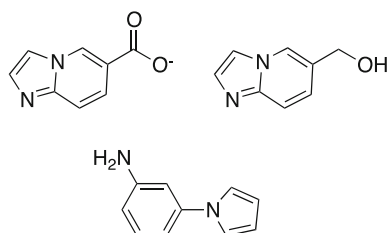


Fig. 9 Outliers in fragment affinity prediction. In clockwise direction from the *top left*, the fragments are frag.aff.15, frag.aff.16 and frag.aff.27 of the SAMPL-3 set

However, the correlation coefficients for affinity prediction are not much improved. For rSIE, r^2 goes from 0.00 to 0.04 after correcting for the outliers. For SIE + FiSH, the r^2 go from 0.00 to 0.02. The MUE is 0.60 and 0.51 kcal/mol for rSIE and SIE + FiSH, respectively, after correcting for outliers.

Outliers of host–guest affinity prediction

Given the good predictions obtained, there are no major outliers for most of the host–guest affinity models. Some of the outliers seen in the scatter plots in Fig. 6 depend on prediction model as well as on whether the outlier analysis refers to absolute or only correlating binding affinities. For example, in the case of host-1, the standard SIE model #94 indicates two outliers, guest-6 and guest-7, with absolute binding affinities underestimated by more than 2 kcal/mol. However, the rescaled SIE model (retrospective), which changes the correlation slope and the spread of prediction data but not the degree of correlation, shows that guest-6 is well-predicted. Although guest-7 pose docked into the ionized host-1 interacts well by traversing the entire central hole

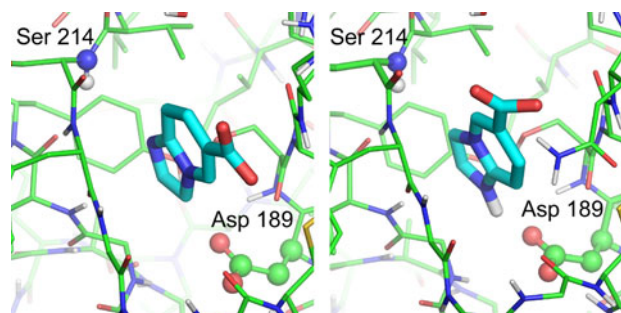


Fig. 10 Predicted binding mode of an outlier, frag.aff.15. *Left panel* Unprotonated imidazo nitrogen. *Right panel* Protonated imidazo nitrogen. Trypsin is depicted as thin sticks with the Ser 214 NH and Asp 189 side chain as *ball and sticks*. Protonation of the imidazo group allows an ion-pair interaction to be formed with the Asp 189 side chain

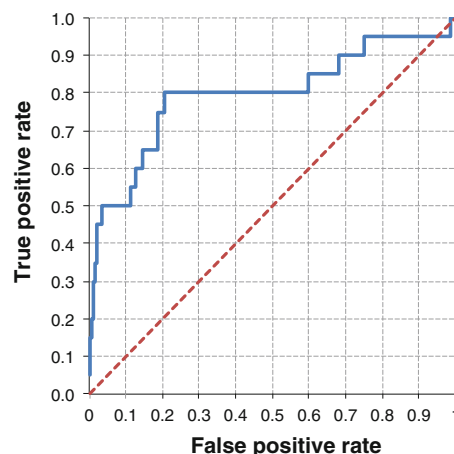


Fig. 11 ROC for trypsin virtual screening after correcting for outliers. The SIE + FiSH scoring function was used

and engages both the amine and amide protons in H-bonds with the host (Fig. 8), its binding affinity is still underestimated. Interestingly, guest-7 was one of the few ligands affected by the protonation state of the host-1, the case of neutral host-1 leading to an improved prediction (model #101, Fig. 6; Table S1) despite the fact that the docked pose does not form any direct H-bonds nor does it fully cross from one face to the other of the neutral host-1 (Fig. 8). This serves to remind us about the delicate balance between interaction and desolvation, and the extent to which scoring functions can accurately account for that balance.

The outlier analysis also points out that system-specific rescaling of scoring functions for predicting absolute binding affinities has to be done carefully and not fully rely on the mathematical global optimum set of parameters but also consider the physical relevance of the system. In this study, the global optimum during training was found to correspond to the neutral host-1, with the resulting rescaled scoring function overshooting the predicted absolute binding affinities of the blind set towards overestimation, with guest-1 and guest-3 as significant outliers (model #99, Fig. 6). The more physically sound ionized state of the host-1 (for the experimental pH of 7.4) produced a rescaled SIE function slightly suboptimal in the training phase, but better performing in the test set (retrospective model in Fig. 6).

Prediction for host-2 and host-3 were reasonable (see the red triangles and yellow squares, respectively, in the scatter plots in Fig. 6). Specifically, the more branched h23-guest-1 due to the *n*-propyl substituent at the N atom of the imidazolyl ring was predicted to fit only partially into the smaller host-2 and hence it has a weaker binding affinity to host-2 (both predicted and experimental) than the more linear h23-guest-2 (the *n*-propyl substituent at the C atom of the imidazolyl ring) which traverses the host-2 central hole (middle panels in Fig. 5). However, same h23-guest-1 and h23-guest-2 were fitted well in the larger host-3 (lower panels in Fig. 5) to which they bind with similar affinities. Although absolute binding affinities to host-3 are slightly underestimated by the standard SIE function (e.g., model #94) by 1–2 kcal/mol, they are well predicted by the rescaled SIE function (retrospective model in Fig. 6).

As mentioned earlier, the use of the SIE + FiSH scoring function provided underestimated absolute binding affinities for all host–guest complexes, with guest-7 of host-1 as major outlier underestimated by 10 kcal/mol (model #98). Rescaling of the SIE + FiSH function did not change the modest low correlation with experimental data, but reduced significantly absolute errors in most outliers and provided more balanced absolute predictions (model #100, Fig. 6). The largest outlier with the rescaled SIE function is guest-1 for host-1, which was predicted as a racemic mixture from the binding poses of the enantiomers (Fig. 12).

Assessment of general performance of SIE

It is informative to position the current prospective SIE predictions of binding affinity at SAMPL-3 in the context of the general performance of the SIE function during training and various tests available thus far.

Training

Standard SIE parametrization (see “Methods” section) was previously derived on a protein–ligand dataset consisting of 99 complexes from 11 diverse protein targets, each comprising a short congeneric series of ligands with known binding affinities curated from the literature and co-crystal structures solved at high-resolution [15]. A training performance characterized by an MUE of 1.34 kcal/mol and an r^2 of 0.65 (Fig. 13a) was obtained while maintaining the physical meaning and interpretability of the optimal parameters. In particular, the fitted optimal solute dielectric falls within the range of 2–4 in agreement with refractive index measurements of protein powders, and there is a scaling down of the potential energy plus solvation by about 90% likely reflecting the compensation exerted by the configurational entropy loss arising from narrowing of the energy wells in the complex versus the free state [57, 58].

CSAR

The most extensive testing of the SIE function was done recently in the Community Structure-Activity Resource (CSAR) scoring challenge consisting of high-resolution co-crystal structures for 343 protein–ligand complexes with high-quality binding affinity data and high diversity with respect to protein targets [32–34]. While the dataset

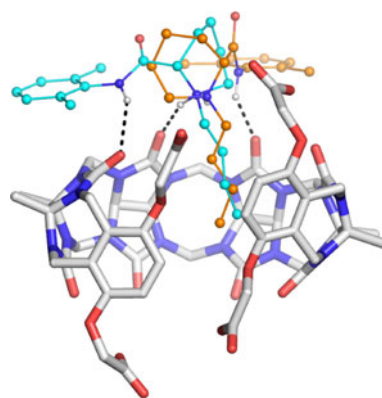


Fig. 12 Binding of the two enantiomers of guest-1 to host-1. The example shown is based on the SIE-selected top-ranked pose and an ionized state of host-1. The C atoms of the guest (ball-and-stick models) are colored in *green* and *cyan* for the *R* and *S* enantiomers, respectively. H-bonds are indicated as *black dashed line*. Only polar H atoms are shown for clarity

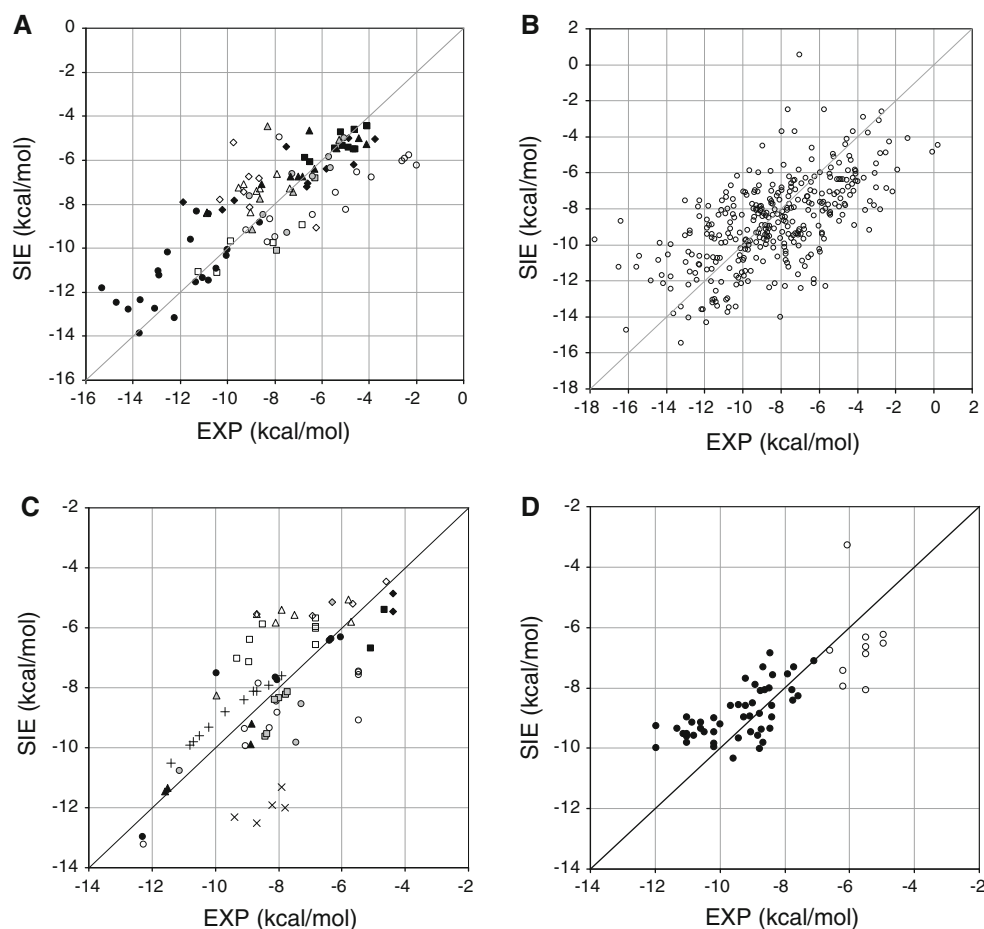


Fig. 13 Past performance of the SIE function on various datasets. **a** Calibration dataset consisting of 99 protein–ligand complexes [15]. HIV protease (*filled circle*); trypsin (*open circle*); lysozyme (*filled square*); thrombin (*open square*); neuraminidase (*filled diamond*); elastase (*open diamond*); triosephosphate isomerase (*filled triangle*); L-arabinose binding protein (*open triangle*); protein tyrosine phosphatase 1B (*grey-shaded circle*); glutathione transferase (*grey-shaded square*); streptavidin (*grey-shaded diamond*). **b** Testing on the CSAR-NRC-HiQ data set ($N = 343$) of the CSAR-2010 scoring exercise [34]. **c** Published applications to date including both SIE predictions and actual binding affinities. References: [18] $N = 7$ (*filled circle*); [23] $N = 12$ with 5 limiting values (*open circle*); [20] $N = 2$ (*filled*

square); [24] $N = 8$ with 4 limiting values and IC50 data used (*open square*); [21] $N = 2$ (*filled diamond*); [26] $N = 4$ (*open diamond*); [22] $N = 4$ (*filled triangle*); [19] $N = 7$ (*open triangle*); [25] $N = 11$ (*plus sign*); [27] $N = 3$ (*grey-shaded circle*); [28] $N = 6$ (*grey-shaded square*) with IC50 data used; [29] $N = 5$ (*multiplication sign*); [30] $N = 1$ (*grey-shaded triangle*); [31] $N = 1$ (*grey-shaded diamond*) with IC50 data used (*multiplication sign*). **d** Blind testing on the SAMPL-1 affinity prediction JNK3 dataset of 59 complexes [35]. *Closed symbols (filled circle)* correspond to active inhibitors with co-crystal structures, and *open symbols (open circle)* correspond to weak binders computationally docked into the kinase active site

resembles the SIE calibration dataset of 99 protein–ligand complexes in terms of target diversity and curation quality, there is no single entry in the CSAR-2010 dataset that was present in the SIE calibration dataset, albeit some protein targets were represented in both data sets. The previously calibrated standard SIE parametrization predicted absolute binding affinities for the highly curated CSAR-NRC-HiQ data set version well in the range of the experimental values, with an MUE of 1.98 kcal/mol and an r^2 of 0.38 (Fig. 13b). SIE predictions were found to be sensitive to the assignment of protonation and tautomeric states in the complex, and the treatment of metal ions near the protein–ligand interface. These structural preparation steps were

critical for accurate testing of the SIE performance. Retraining and testing of SIE parameters on two predefined halves of CSAR-NRC-HiQ led to only marginal further improvements to an MUE of 1.83 kcal/mol and an r^2 of 0.43, with modest change in the optimal values of SIE parameters.

Published studies

The SIE function has also been applied retrospectively as well as prospectively in several other independent laboratories that have reported SIE predictions versus actual binding affinities [18–31]. Collectively, these data indicate

an MUE of 1.30 kcal/mol and an r^2 of 0.47 between the predicted and actual absolute binding affinities (Fig. 13c). As in the case of the CSAR-NRC-HiQ data set, these applications reiterate that the SIE approach returns predicted protein–ligand binding affinities well within the range of experimental measurements. The degree of scatter is comparable to that observed in the original calibration and in the CSAR testing, suggesting that the SIE parameters were not over-fitted to the training set.

SAMPL-1

Since the most objective way to evaluate computational methods is via blind tests, SIE was a participating method in the SAMPL-1 experiment organized by OpenEye, Inc. in early 2008 [35]. In SAMPL-1, we tested prospectively the standard SIE parametrization for protein–ligand binding affinity prediction on the Jun kinase 3 (JNK3) data set, a target class not used for SIE calibration. This data set consisted of 49 diverse JNK3 inhibitors from 12 classes, each with its own co-crystal structure with the kinase, plus 10 models of known “inactive” ligands (in fact weakly active ligands) docked in duplicated enzyme structures. The SIE function achieved reasonable prospective predictions for the JNK3 dataset of 49 actives, with an MUE of 0.92 kcal/mol and an r^2 of 0.36. Again, it became apparent that SIE can estimate absolute binding affinities, with predicted values spanning the same range as the actual ones (Fig. 13d). The 10 measured inactives were separated reasonably well from the actives, leading to an increase in r^2 to 0.54 over all 59 ligands.

SAMPL-3

As described in this paper, the SIE function returned encouraging prospective predictions when tested on the trypsin-fragment and host–guest blind data sets from the SAMPL-3 experiment. (Figure 6 on host–guest and Fig. 3 on trypsin-fragment data sets.) SAMPL-3 was a useful experiment because it tested the applicability domain of the method with challenging systems like fragment-sized weak-affinity ligands binding to an enzyme, and small host–guest systems exhibiting appreciable binding affinities. Additionally, these predictions had to be made not on solid, experimentally determined, binding modes as in CSAR and SAMPL-1, but on computationally docked binding modes. We also tested more extensively SIE rescaled specifically for the system under investigation, as well as we tested for the first time SIE + FiSH, a scoring function that incorporates our latest solvation model, FiSH [42], into SIE. We found that for the trypsin-fragment data set, rescaling of the SIE parameters was necessary to improve prediction of absolute binding affinities (MUE of

0.81 kcal/mol), which were systematically overestimated by the standard SIE parametrization (MUE of 2.24 kcal/mol). We note that even in the original SIE training set, the trypsin-ligand subset was also overestimated (Fig. 13). For the host–guest system, the MUE of about 1.16 kcal/mol and r^2 of 0.52 kcal/mol were hardly affected by SIE rescaling, but the correlation slope became closer to 1 after rescaling due to a larger entropy-related factor, in agreement with other studies [37, 38]. This suggests that in certain cases, possibly for fragment-sized ligands and other small molecular systems, the SIE function may need to be retrained for the system under investigation if data are available. A recent study also reports improved predictions with rescaled SIE parameters for protein–ligand systems, although the system-specific sets of SIE parameters were not validated on external sets [64]. Encouragingly in the case the trypsin-fragment data set, the SIE + FiSH scoring function outperformed the standard SIE scoring function in terms of absolute predictions (MUE of 2.24 kcal/mol for standard SIE versus 0.98 kcal/mol for standard SIE + FiSH). However, testing of the SIE + FiSH scoring function on more systems is required in order to confirm its general advantage.

Virtual screening

The compromise between speed and accuracy makes SIE a suitable scoring function for ranking compound libraries in virtual screening (VS) applications. Previously, SIE was tested for VS enrichment against estrogen receptor (ER) and thymidine kinase (TK) showing the ability of SIE to recover true hits in a collection of decoys [15]. While the ER set is considered an easier test, the TK set is more challenging partly due to weaker binding affinities for the true binders. The SIE function was able to recover all true positives within the top 10% of the ranked dataset, and half of them within top 1%. The SIE was clearly superior to simpler functions, e.g., buried surface area that describes only non-polar effects and ranked all TK true binders near the bottom of the list. In the blinded VS experiment of SAMPL-3, SIE showed a strong performance on the trypsin-fragment dataset of over 500 ligands, significantly enriching in the 20 true-active fragments with an AUC value of 0.70. A promising retrospective result is that the SIE + FiSH function improves the enrichment (AUC of 0.73) in this VS data set.

Docking

Although SAMPL-3 did not explicitly test docking methods it is clear that success or failure in virtual screening or affinity prediction is highly dependent on the quality of the predicted poses that are scored. For this purpose we opted

to go with an exhaustive docker, Wilma, that thoroughly samples bound conformations rather than a stochastic docker with uncertain convergence properties. The rather fine search grid used (0.5 Å) combined with thorough sampling of ligand rotamers using OMEGA gives us some confidence that the native pose was visited during the search procedure. The scoring function used for docking is Physics-based and mimics the major components of a typical force-field calculation, albeit with empirically modified weights for the various terms. The net effect is that the top poses selected by Wilma will most likely be low-energy poses as well when rescored with our SIE function. In fact, this is what we observed by noting that the affinity predictions using the Top-Wilma pose were comparable to those using the Top-SIE pose. This is no mean feat given that the Wilma docking function is several orders of magnitude faster to compute than the SIE function.

Given the speed of Wilma scoring, it is tempting to use Wilma scoring alone for virtual screening applications. However, in our experience virtual screening ranking based on Wilma scores alone are not as reliable as that obtained after rescoring with the SIE function (data not shown). This is probably because in docking a given molecule many terms cancel out when comparing the energetics of one pose versus another. There is much less cancellation when comparing the affinity of one molecule versus another with a different molecular structure. Hence, a function optimized for docking may not properly capture key components necessary for accurate affinity prediction across different molecules. We find that the use of Wilma for docking and SIE for scoring achieves a cost-effective balance between speed and accuracy for both virtual screening and affinity prediction.

Conclusions

The performance of the SIE scoring function in this blind test is consistent with past experience in its application to a number of targets. The combination of SIE with an exhaustive docker such as Wilma affords a rapid cost-effective virtual screening platform that can provide not just a ranking of compounds but estimates of binding affinity as well. The sampling thoroughness afforded by Wilma may in fact be instrumental for the relatively good results obtained in virtual screening and affinity prediction. Nevertheless, the goal of consistently achieving sub-2 kcal/mol accuracy in relative binding free energies remains a challenge, as seen in poor correlations when the dynamic ranges of the actual binding affinities are narrow. However, it is encouraging that the inclusion of more Physics in the

model, e.g., SIE + FiSH, can improve the quality of the predictions.

Acknowledgments This is National Research Council of Canada publication number 53158.

References

- Reddy MR, Erion MD (2001) Free energy calculations in rational drug design. Springer, Berlin
- 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(2): 150–160
- 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
- Gilson MK, Zhou HX (2007) Calculation of protein–ligand binding affinities. *Annu Rev Biophys Biomol Struct* 36(1):21–42
- 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
- Wang R, Lu Y, Fang X, Wang S (2004) An extensive test of 14 scoring functions using the PDB bind refined set of 800 protein–ligand complexes. *J Chem Inf Comput Sci* 44:2114–2125
- 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(20):5912–5931
- 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(S1):S7–S26
- 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(6): 1568–1580
- Zou X, Sun Y, Kuntz ID (1999) Inclusion of solvation in ligand binding free energy calculations using the generalized-born model. *J Am Chem Soc* 121:8033–8043
- 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(12):889–897
- Kuhn B, Gerber P, Schulz-Gasch T, Stahl M (2005) Validation and use of the MM-PBSA approach for drug discovery. *J Med Chem* 48(12):4040–4048
- 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(2):238–250
- 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(10):3159–3165
- Naïm M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, Siddiqi I, Drabik P, Sulea T, Bayly C, 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(1):122–133

16. Cui Q, Sulea T, Schrag JD, Munger C, Hung M-N, Naïm 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(4):787–802
17. Sulea T, Purisima EO (2011) The solvated interaction energy (SIE) method for scoring binding affinities. In: Baron R (ed) *Methods in molecular biology, computer-aided drug design*. Humana Press (Springer Publishing Group) (in press)
18. Wang YT, Su ZY, Hsieh CH, Chen CL (2009) Predictions of binding for dopamine D2 receptor antagonists by the SIE method. *J Chem Inf Model* 49(10):2369–2375
19. Mishra NK, Kríz Z, Wimmerová M, Koca J (2010) Recognition of selected monosaccharides by *Pseudomonas aeruginosa* lectin II analyzed by molecular dynamics and free energy calculations. *Carbohydr Res* 345(10):1432–1441
20. Rodriguez-Granillo A, Sedlak E, Wittung-Stafshede P (2008) Stability and ATP binding of the nucleotide-binding domain of the Wilson disease protein: effect of the common H1069Q mutation. *J Mol Biol* 383(5):1097–1111
21. Wei C, Mei Y, Zhang D (2010) Theoretical study on the HIV-1 integrase-SCITEP complex based on polarized force fields. *Chem Phys Lett* 495(1–3):121–124
22. Lecaille F, Chowdhury S, Purisima E, Brümme D, Lalmanach G (2007) The S2 subsites of cathepsins K and L and their contribution to collagen degradation. *Protein Sci* 16(4):662–670
23. Nguyen M, Marcellus RC, Roulston A, Watson M, Serfass L, Murthy Madiraju SR, Goulet D, Viallet J, Belec L, Billot X, Acoca S, Purisima E, Wiegman A, Cluse L, Johnstone RW, Beauparlant P, Shore GC (2007) Small molecule obatoclox (GX15-070) antagonizes MCL-1 and overcomes MCL-1-mediated resistance to apoptosis. *Proc Natl Acad Sci USA* 104(49):19512–19517
24. Okamoto M, Takayama K, Shimizu T, Muroya A, Furuya T (2010) Structure–activity relationship of novel DAPK inhibitors identified by structure-based virtual screening. *Bioorg Med Chem* 18(7):2728–2734
25. Yang B, Hamza A, Chen G, Wang Y, Zhan C-G (2010) Computational determination of binding structures and free energies of phosphodiesterase-2 with benzo[1,4]diazepin-2-one derivatives. *J Phys Chem B* 114(48):16020–16028
26. Wimmerová M, Mishra N, Pokorn M, Koca J (2009) Importance of oligomerisation on *Pseudomonas aeruginosa* lectin-II binding affinity. In silico and in vitro mutagenesis. *J Mol Model* 15(6):673–679
27. Hamza A, Zhao X, Tong M, Tai H-H, Zhan C-G (2011) Novel human mPGES-1 inhibitors identified through structure-based virtual screening. *Bioorg Med Chem* 19(20):6077–6086
28. Coluccia A, Sabbadin D, Brancale A (2011) Molecular modelling studies on arylthioindoles as potent inhibitors of tubulin polymerization. *Eur J Med Chem* 46(8):3519–3525
29. Anisimov VM, Cavasotto CN (2011) Quantum mechanical binding free energy calculation for phosphopeptide inhibitors of the Lck SH2 domain. *J Comput Chem* 32(10):2254–2263
30. Hartono YD, Lee AN, Lee-Huang S, Zhang D (2011) Computational study of bindings of HL9, a nonapeptide fragment of human lysozyme, to HIV-1 fusion protein gp41. *Bioorg Med Chem Lett* 21(6):1607–1611
31. Duque MD, Ma C, Torres E, Wang J, Naesens L, Juarez-Jimenez J, Camps P, Luque FJ, DeGrado WF, Lamb RA, Pinto LH, Vasquez S (2011) Exploring the size limit of templates for inhibitors of the M2 ion channel of influenza A virus. *J Med Chem* 54(8):2646–2657
32. 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(9):2036–2046
33. Smith RD, Dunbar JB, Ung PM-U, Esposito EX, Yang C-Y, Wang S, Carlson HA (2011) CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. *J Chem Inf Model* 51(9):2115–2131
34. 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
35. Moghaddam S (2008) SAMPL1 at first glance. CUP IX meeting, Santa Fe, NM, 19 March 2008. http://eyesopen.com/2008_cup_presentations/CUP9_Skillman.pdf. Accessed 1 Oct 2011
36. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* 6(3):211–219
37. 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
38. Moghaddam S, Inoue Y, Gilson MK (2009) Host–guest complexes with protein–ligand-like affinities: computational analysis and design. *J Am Chem Soc* 131(11):4012–4021
39. McInnes C (2007) Virtual screening strategies in drug discovery. *Curr Opin Chem Biol* 11(5):494–502
40. Fennell CJ, Kehoe CW, Dill KA (2011) Modeling aqueous solvation with semi-explicit assembly. *Proc Natl Acad Sci USA* 108(8):3234–3239
41. Fennell CJ, Kehoe C, Dill KA (2009) Oil/water transfer is partly driven by molecular shape, not just size. *J Am Chem Soc* 132(1):234–240
42. 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
43. 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
44. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the protein databank and Cambridge structural database. *J Chem Inf Model* 50(4):572–584
45. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, 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
46. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The amber biomolecular simulation programs. *J Comput Chem* 26(16):1668–1688
47. 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
48. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 97:10269–10280
49. Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J Am Chem Soc* 115:9620–9631
50. 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
51. 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

52. Purisima EO (1998) Fast summation boundary element method for calculating solvation free energies of macromolecules. *J Comput Chem* 19(13):1494–1504
53. Purisima EO, Nilar SH (1995) A simple yet accurate boundary element method for continuum dielectric calculations. *J Comput Chem* 16:681–689
54. Chan SL, Purisima EO (1998) A new tetrahedral tessellation scheme for isosurface generation. *Comput Graph* 22(1):83–90
55. Chan SL, Purisima EO (1998) Molecular surface generation using marching tetrahedra. *J Comput Chem* 19(11):1268–1277
56. Bhat S, Purisima EO (2006) Molecular surface generation using a variable-radius solvent probe. *Proteins Struct Funct Bioinf* 62(1):244–261
57. Chang CE, Gilson MK (2004) Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. *J Am Chem Soc* 126(40):13156–13164
58. Chen W, Chang CE, Gilson MK (2004) Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. *Biophys J* 87(5):3035–3049
59. Ma D, Zavalij PY, Isaacs L (2010) Acyclic cucurbit[n]uril congeners are high affinity hosts. *J Org Chem* 75(14):4786–4795
60. Mobley DL, Barber AE, Fennell CJ, Dill KA (2008) Charge asymmetries in hydration of polar solutes. *J Phys Chem B* 112(8):2405–2414
61. Purisima EO, Sulea T (2009) Restoring charge asymmetry in continuum electrostatics calculations of hydration free energies. *J Phys Chem B* 113(24):8206–8209
62. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935
63. Floris F, Tomasi J (1989) Evaluation of the dispersion contribution to the solvation energy. A simple computational model in the continuum approximation. *J Comput Chem* 10(5):616–627
64. Lill MA, Thompson JJ (2011) Solvent interaction energy calculations on molecular dynamics trajectories: increasing the efficiency using systematic frame selection. *J Chem Inf Model* 51(10):2680–2689