Binding affinities in the SAMPL3 trypsin and host–guest blind tests estimated with the MM/PBSA and LIE methods

Paulius Mikulskis [1], Samuel Genheden [1*], Patrik Rydberg [2], Lars Sandberg[3], Lars Olsen[2], Ulf Ryde[1]

[1] Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124,
SE-221 00 Lund, Sweden

[2] Biostructural Research, Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark

[3] Innovative Medicines, AstraZeneca R&D, SE-151 85 Södertälje, Sweden

Correspondence to Samuel Genheden, E-mail: Samuel.Genheden@teokem.lu.se,
Tel: +46 – 46 2224915, Fax: +46 – 46 2228648

2013-01-29

Abstract

We have estimated affinities for the binding of 34 ligands to trypsin and nine guest molecules to three different hosts in the SAMPL3 blind challenge, using the MM/PBSA, MM/GBSA, LIE, continuum LIE, and Glide score methods. For the trypsin challenge, none of the methods were able to accurately predict the experimental results. For the MM/GB(PB)SA and LIE methods, the rankings were essentially random and the mean absolute deviations were much worse than a null hypothesis giving the same affinity to all ligand. Glide scoring gave a Kendall's τ index better than random, but the ranking is still only mediocre, τ = 0.2. However, the range of affinities is small and most of the pairs of ligands have an experimental affinity difference that is not statistically significant. Removing those pairs improves the ranking metric to 0.4-1.0 for all methods except CLIE. Half of the trypsin ligands were non-binders according to the binding assay. The LIE methods could not separate the inactive ligands from the active ones better than a random guess, whereas MM/GBSA and MM/PBSA were slightly better than random (area under the receiver-operating-characteristic curve, AUC = 0.65–0.68), and Glide scoring was even better (AUC = 0.79). For the first host, MM/GBSA and MM/PBSA reproduce the experimental ranking fairly good, with τ = 0.6 and 0.5, respectively, whereas the Glide scoring was considerably worse, with a τ = 0.4, highlighting that the success of the methods is system-dependent.

Key Words: Binding affinity, MM/GBSA, MM/PBSA, linear interaction energy, docking, Glide.

Introduction

The estimation of the free energy for the binding of a small molecule to a macromolecule is one of the greatest challenges in computational chemistry [1]. For example, if the binding affinity of a drug candidate to its biomolecular target could be accurately predicted by calculations, enormous amounts of money could be saved by pharmaceutical companies. Likewise, the catalytic power of enzymes can be formulated as the preferential binding of transition states before substrates and products. Unfortunately, the progress in calculating binding affinities has been mediocre, partly owing to the fact that biomacromolecules are large and complicated systems. Therefore, a growing interest has been directed towards host–guest systems, which are organic complexes that bind specific ligands. The binding is dictated by the same intermolecular forces as in a macromolecule–ligand complex, but because of their smaller size, there is less phase space to sample [2]. Hence, they are excellent tests cases for theoretical methods. In addition, they have interesting applications themselves, e.g. for the transport of drug molecules.

Much effort has been spent on the development of methods to estimate binding affinities, ranging from simple statistical scoring functions to strict physical methods based on statistical mechanics [1]. Free-energy perturbation and thermodynamic integration are in principle exact methods, but they are very time-consuming, because they require extensive sampling of unphysical intermediate states. Therefore, several more approximate methods have been developed that are also based on molecular dynamics (MD) or Monte Carlo sampling, but only of the physical end states (the free ligand, the free receptor, and their complex). Examples of such methods are PDLD/s-LRA/β (semi-macroscopic protein-dipoles Langevin-dipoles method within a linear-response approximation) [3,4], LIE [5,6,7] (linear interaction energy), and MM/PB(GB)SA [8,9] (molecular mechanics with Poisson–Boltzmann (or generalised Born) and surface-area solvation). These methods have become popular tools to estimate the binding free energy of molecular complexes. For example, between 2006 and 2010, ISI Web of Knowledge includes 283 articles with MM/GBSA or MM/PBSA as the topic and 84 articles with LIE.

In a series of publications we have evaluated the accuracy, precision, and stability of the MM/PB(GB)SA method to predict protein–ligand affinities: We have developed an approach that gives MM/GBSA energies with a statistical precision of 1 kJ/mol [10], ensuring that the results are reproducible [11]. Moreover, we have improved the entropy term [12] and tested the effect of changing or improving the electrostatic [13,14], the polar [15,16] and the non-polar solvation energies [17,18], or even the whole molecular-mechanics energy term [19]. Recently, we have evaluated the precision of the LIE method also and compared with the results of the MM/GBSA method [20]. In addition, we have investigated the effect of the force field and solvation model on the prediction of host–guest binding affinities [21]. Consequently, we have gained much experience with these methods.

Naturally, there is a great interest in comparing the accuracy different ligand-binding methods. Unfortunately, such comparisons between MM/PB(GB)SA and LIE are quite few and the results are varying: For the binding of biotin analogous to avidin, MM/PBSA and MM/GBSA outperformed LIE [20,22], whereas for acetylcholinesterase huprine inhibitors, LIE gave better results than MM/PBSA [23]. For the binding of eight hydroxamate inhibitors to gelatinase-A and the binding of fragment B of protein A to the Fc domain of immunoglobin G, the two methods showed a similar performance [24,25]. Finally, for the binding of primary alcohols to cyclodextrin, LIE was more accurate than MM/GBSA and MM/PBSA, whereas for the binding of guests to cucurbitil[8]uril, LIE was less accurate than MM/GBSA and MM/PBSA [21]. Apparently, the performance of the two methods depends on the model system, so more tests are needed.

Another problem with such comparisons is that if the results are known beforehand, it might be tempting to rerun calculations that gave poor results. If the standard deviation is large enough (as it often is for MM/PBSA [13), this will nearly always improve the results. Of course, this problem can be avoided by increasing the precision of the method [14,15,20]. However, a blind test, in which the experimental results are not known when the calculations are run, is even stronger. The SAMPL challenges, organised by OpenEye Scientific Software, offer a unique opportunity to make an assessment of various methods [26,27]. In this article, we present our MM/PBSA, MM/GBSA, and LIE results for two of the SAMPL3 challenges, which involve the binding of 34 ligands to trypsin [28] and nine guests to three different hosts. It should be noted that both MM/PB(GB)SA and LIE are methods to estimate only ligand-binding affinities, but these tests involve also the decision of the protonation states and the binding modes of the ligands. This has to be done with other software (or by chemical intuition) and the choices will strongly affect the results. On the other hand, it has allowed us to compare also with the results obtained with the Glide docking software [29].

Methods

Preparation of trypsin and the trypsin ligands. The protonation states of the trypsin ligands (L1–L34 in Figure 1) were estimated by the Epik tool [30]. In two ambiguous cases (L15 and L16), pK_a calculations were performed using the pK_a-prediction module in Jaguar, which combines density functional theory calculations at the B3LYP level [31,32,33] with empirical corrections [34,35]. The final protonation state of the ligands, selected to apply to a pH of 7, is shown in Figure 1. The ligands were docked into the frag.aff.tryp1.pdb trypsin crystal structure provided by the organisers. Protonation states of amino acid side chains in the trypsin structure were assigned using the protein preparation wizard in Maestro [36] and docking was performed with the Glide program [29,37] using the standard-precision scoring function.

The MD simulations were based on the 1hj9 crystal structure [38] and the protein was prepared as follows. Because we intended to perform LIE calculations, it is necessary to neutralise the protein. Therefore, all Asp and Glu residues were protonated and all Lys and Arg residues were deprotonated with the following five exceptions: The two Glu residues that bind the Ca^{2+} ion (Glu-70 and 80) were left deprotonated (the other Ca^{2+} ligands are the back-bone O atom of Asn-72 and Val-75, and two water molecules). In addition, Asp-194 and Lys-107, which make ionic pairs with the charged N- and C-terminal residues, respectively, were left in their charged state. Asp-189, which is at the bottom of the S1 binding pocket, possibly interacting with a positive charge on the ligand, was also left deprotonated. Finally, His-57 in the catalytic triad was doubly protonated, whereas the other two His residues were protonated on the NE2 atom.

The protein was modelled by the Amber99SB force field [39] and the ligands were described with the general Amber force field (GAFF) [40], using charges calculated with the RESP method [41], based on quantum mechanical calculations of the electrostatic potential at the Hartree–Fock level with the 6-31G* basis set and points sampled with the Merz–Kollman scheme [42]. The Ca^{2+} ion was described with a non-bonded potential, using the formal +2 charge and Lennard-Jones parameters of 1.60 Å and 0.42 kJ/mol (from the Amber parm91.dat file). This gave Ca–O distances of ~2.4 Å in the MD simulations. Each protein–ligand complex was immersed in a truncated octahedral box of TIP3P water molecules [43] that extended at least 10 Å outside the protein.

Preparation of host–guest complexes. The guests for host1 (G1–G7 in Figure 2) were docked into the host with the Glide program [37], using a receptor grid large enough to include the full host1 structure and the Glide extra-precision docking mode with default options [29**]**. The guests for host2 and host3 (G8 and G9 in Figure 2) were manually placed parallel to and roughly centred in the symmetric host so that the four nitrogen atoms were lining the rim of the host (see Figure S1). The protonation states of the guests were assigned manually as is shown in Figure 2. All guests and hosts were described using GAFF [13] with AM1-BCC charges [44]. The N atoms in G8 and G9 were described using the same atom types. The complexes were immersed in pre-equilibrated, truncated octahedral boxes of TIP3P water molecules [13] extending at least 10 Å outside the solute.

Simulation protocol. All MD simulations were run by either the sander or pmemd modules in Amber 11 [45]. The temperature was kept at 300 K using Langevin dynamics [46] with a collision frequency of 2.0 ps^{-1}. The pressure was kept at 1 atm using a weak-coupling approach [47] with isotropic position rescaling and relaxation time of 1 ps. The long-range electrostatics was treated by particle-mesh Ewald summation [48] with a fourth-order B-spline interpolation and a tolerance of 10^{-5}. The non-bonded cutoff was 8 Å and the non-bonded pair list was updated every 50 fs. The MD time step was 2 fs and the SHAKE algorithm [49] was used to constrain the lengths of bonds involving hydrogen atoms.

The trypsin complexes and all free ligands were simulated in the following way: The system was minimised with restraints on all non-hydrogen atoms except those in water molecules. Then, 40 independent simulations were initiated by assigning different initial velocities. Each of these simulations, were further equilibrated for 20 ps in the *NPT* ensemble using the same restraints, followed by a 1 ns unrestrained equilibration and a 200 ps production run, in which snapshots where extracted every fifth ps.

The host–guest complexes were simulated in the following way: First, the complex was minimised with restraints on all non-hydrogen atoms except those in water molecules, followed by a 20 ps restrained MD simulation in the *NPT* ensemble, and a 5 ns unrestrained equilibration. After this equilibration, 20 independent simulations were initiated by assigning different initial starting velocities. Each of the simulations was further equilibrated for 50 ps in the *NPT* ensemble, before a 200 ps production run was performed, in which snapshots were extracted every fifth ps.

MM/PB(GB)SA calculations. MM/PB(GB)SA binding energies were calculated according to

$$\Delta G = \langle G(\mathrm{RL}) - G(\mathrm{R}) - G(\mathrm{L}) \rangle_{\mathrm{RL}} \qquad (1)$$

where RL is the complex, R is the receptor (either trypsin or the hosts), and L is the ligand (either the trypsin ligands or the guests). The brackets indicate an ensemble average over a MD simulation of RL. Each free-energy term was calculated from:

$$G = E_{\mathrm{ele}} + E_{\mathrm{vdw}} + G_{\mathrm{solv}} + G_{\mathrm{np}} - TS \qquad (2)$$

The E_{ele} and E_{vdW} terms in Eqn. 2 represent electrostatic and van der Waals energies and they were calculated with Amber [40] with all water molecules stripped off and without any periodic boundary conditions, but with an infinite cutoff. The polar solvation energy, G_{solv}, was estimated by either a generalised Born (GB) model or by solving the Poisson–Boltzmann (PB) equation. We used the GB model I of Onufriev, Bashford and Case with $\alpha = 0.8$, $\beta = 0$, and $\gamma = 2.91$ (GB$^{\mathrm{OBCI}}$) [50], or the PB solver in Amber10 with a grid spacing of 0.5 Å and a probe radius of 1.4 Å. The non-polar part of the solvation energy, G_{np}, was estimated from the solvent-accessible surface area (SASA) according to $G_{\mathrm{np}} = \gamma\mathrm{SASA} + b$ with $\gamma = 0.0227$ kJ/mol/Å^2 and $b = 3.85$ kJ/mol. The entropy (*S* in Eqn. 2) was estimated by a normal-mode analysis of the harmonic frequencies, calculated at the MM level. For the protein calculations, we used our recent modification of the MM/PBSA approach with improved precision [12]: All residues more than 12 Å from any atom in the ligand were deleted and the remaining atoms were minimised, keeping all residues more than 8 Å from ligand fixed (including all water molecules), to ensure that the geometry is as close as possible to the original structure. In the frequency calculations, the fixed buffer region was omitted. For the host calculations, we used the standard method, using a minimisation with a distance-dependent dielectric constant, ε = 4 *r*.

LIE and CLIE calculations. LIE energies were estimated according to

$$\Delta G = \beta\left(\langle E_{\mathrm{ele}}^{\mathrm{L\text{-}S}}\rangle_{\mathrm{RL}} - \langle E_{\mathrm{ele}}^{\mathrm{L\text{-}S}}\rangle_{\mathrm{L}}\right) + \alpha\left(\langle E_{\mathrm{vdw}}^{\mathrm{L\text{-}S}}\rangle_{\mathrm{RL}} - \langle E_{\mathrm{vdw}}^{\mathrm{L\text{-}S}}\rangle_{\mathrm{L}}\right) \tag{3}$$

where $E_{\mathrm{ele}}^{\mathrm{L\text{-}S}}$ and $E_{\mathrm{vdw}}^{\mathrm{L\text{-}S}}$ are the electrostatic and van der Waals intermolecular interaction energies between the ligand and the surroundings, *α* and *β* are two parameters, and the angle brackets indicate ensemble averages over simulations of either the free ligand or the complex, as indicated by the subscripts. In one set of calculations, the *α* parameter was set to the default value of 0.18 for all ligands and *β* was varied depending on nature of ligand (0.5 for charged ligands, 0.43 for neutral ligands without any hydroxyl group, and 0.37 for neutral ligands with a single hydroxyl group[6]). In a second set of calculations, we optimised the *α* parameter with respect to MADtr, keeping the same *β* values as in the first set.

We have also used a continuum variant of LIE (CLIE) [51]. In these calculations, the two $E_{\mathrm{ele}}^{\mathrm{L\text{-}S}}$ terms were estimated using the $\mathrm{GB}^{\mathrm{OBCI}}$ model after stripping off all water molecules.

Error estimates. All reported uncertainties are standard deviations of the mean, i.e. the standard deviation divided by the square root of the number of samples. The standard deviations were calculated over the 20 or 40 independent simulations, ignoring the uncertainty among the 40 snapshots in each simulation.

The quality of the results, compared to experiments, were quantified using the mean absolute deviation when the systematic error (i.e. the mean signed deviation) has been removed (MADtr), Kendall’s rank correlation coefficient (τ) [52], the correlation coefficient (r^2), and the slope of the best regression line. The uncertainty of these quality metrics was estimated using a parametric bootstrap (using 1000 random samples) [10], utilising the uncertainty of both the experiments and computational predictions.

Result and Discussion

The trypsin challenge

We have estimated the binding affinity of the 34 ligands in the SAMPL3 trypsin challenge. Before the affinities can be estimated, several issues need to be settled. First, a protein structure has to be selected. Two crystal structures were provided by the organisers. However, visual inspection and comparison with other crystal structures indicated that these structures were of a rather poor quality (for example the carbon and nitrogen atoms in His-57 was swapped, no ligand was bound, and no crystal-water molecules were provided, although several previous studies have indicated the importance of water molecules in the active site for ligand binding [53,54]). Therefore, we decided to instead base our MD simulations on the 1hj9 crystal structure, which is at atomic resolution (0.95 Å) and contains a bound aniline molecule (i.e. L3 without the two fluorine atoms).

LIE requires that the complex and free ligand have the same total charge in the simulations. Therefore, we decided to neutralise the protein, as described in the Methods section. To investigate the effect of this neutralisation on the MM/GBSA results, we performed calculations on four ligands, for which experimental affinities have been published (L35–L38, shown in Figure S2) [55,56]. The results of those calculations are shown in Table S1. For ligands L37 and L38, the difference between the calculations with a neutralised protein and with a fully charged protein (0–2 kJ/mol) is not statistically significant. However, for L35 (benzamidine) there is a change of 6 kJ/mol and for L36, the difference is 12 kJ/mol.

Most of the difference for L36 comes from the van der Waals terms and not from electrostatics and polar solvation. These test calculations show that there might be some effect on the absolute MM/GBSA results, although the ranking is preserved between most of the compounds. Because these test calculations did not unambiguously show that neutralisation is detrimental to the results, we decided to only run calculations with the neutralised protein.

Moreover, it is essential to settle the protonation state of the ligands. The protonation state of the ligands were estimated by the Epik software, supplemented by Jaguar calculations for two of the ligands (L15 and L16). Earlier LIE studies have indicated that carboxylate groups bind to trypsin in their protonated (neutral) form [57]. Therefore, we studied the binding of protonated carboxylate groups for ligands L1, L15, and L25, and corrected the binding affinities for the unfavourable protonation of these groups, assuming that the pK_a of the free ligand is equal to that of benzoic acid (4.20 [58], i.e. the calculated binding affinity was reduced by RT ln10 (7 – 4.2) = 16 kJ/mol). However, it should be noted that this correction makes the predictions poorer for all methods, typically rendering them outliers. The final protonation states of the ligands, used in the simulations, are shown in Figure 1. We consider the selected protonation of ligands L3, L15, L16, L27, and L34 somewhat uncertain, and for L13 and L14, the selection is even more ambiguous, because there are two or three sites that can be protonated.

Finally, we docked all ligands into the S1 pocket of trypsin using the Glide software. In general, we took the structures with the best Glide score, but based on available literature and crystal structures [56,59], we required that positively charged groups or, if no such group is present, hydrogen-bond donating groups interact with Asp-189. For L13 and L14, we assumed that the terminal –NH_3 or –OH group binds to Asp-189, rather than the NH group in the ring. For L15, the carboxylate group was assumed to bind to Asp-189. These binding modes are illustrated in Figure 1.

After the submission of the results, crystal structures of complexes with seven of the ligands were published. The RMSD between the docked and crystal poses of the ligands (after a RMSD fit of the protein) was no more than 1.0 Å for all ligands, except L11. This confirms the binding pose for six of the ligands, whereas L11 turned out to bind with the O atom directed towards the solution, rather than inwards into the protein, as the docking suggested (giving a RMSD of 1.97 Å). Therefore, the simulations of this ligand were rerun, using the binding pose observed in the crystal structure. However, this changed the calculated binding affinities by only 0–2 kJ/mol, i.e. within the statistical uncertainty.

We have estimated the binding affinity for the 34 ligands with five different methods: MM/GBSA, MM/PBSA, LIE, CLIE, and Glide scoring. The results are collected in Table 1, together with the corresponding standard errors (not for Glide scoring). For LIE and CLIE, results are presented for both the standard non-polar parameter, α = 0.18, and with this parameter optimised with respect to MADtr (α =0.38 for LIE and α =1.0 for CLIE).

The table also includes experimental data provided by the organisers after the calculations were finished. Experimental data were only provided for 17 of the ligands; for the other 17 ligands, no binding was detected, meaning that the binding free energy is more positive than –17 kJ/mol. The experimental affinity of the ligands for which binding was detected involve a range of only 9 kJ/mol, viz. –18 to –26 kJ/mol. Naturally, such a small range provides a major challenge for computational methods.

Moreover, the use of Kendall's τ rank correlation coefficient (which is simply the number pairs of ligands for which the calculations predict the correct ranking minus the number of pairs with the incorrect ranking, divided by the total number of pairs) becomes problematic, because only 29 of the 136 possible pairs of ligands have a difference in the experimental affinity that is statistically significant at the 95% level (all involving either L6, L12, or L17; the reported experimental uncertainty is 1.7 kJ/mol). If the uncertainties in the computational estimates are also taken into consideration, even fewer pairs have both theoretical and experimental differences that are statistically significant. Therefore, we also calculate a Kendall's τ coefficient, considering only pairs for which both the experimental and predicted differences in affinity are statistically significant at the 95% confidence level. We will denote this metric with τ_{95}.

In Table 1, the quality metrics (MADtr, τ, τ_{95}, r^2, and slope) are computed only for the active compounds. The MM/GBSA predictions show a much larger range of energies (68 kJ/mol) than the experimental binding affinities (9 kJ/mol). The predictions are also in general more negative than the experimental results, by 26 kJ/mol on average. However, for ligands L15, L16, and L27, the predictions are in fact more positive than the experimental results. In addition, L7 is predicted to have an affinity only 7 kJ/mol more negative than the experiments. Naturally, these four ligands stand out in the scatter plot shown in Figure 3a (squares). It is also clear that MM/GBSA fails to recognize the three ligands with the strongest binding (L6, L12, and L17). The MADtr of MM/GBSA is 16 kJ/mol, which is much larger than the null hypothesis that all ligands have the same affinity, which gives a MADtr of 2 kJ/mol. This mainly reflects the large range of predicted affinities (it could be reduced by scaling down all estimated affinities). The ranking of the ligands as quantified by Kendall's τ = 0.10 is almost random and there is no correlation between the calculated and measured affinities (r^2 = 0.01). However, only 20 of the pairs have predicted affinity differences that are significant and these give τ_{95} = 0.70, which shows that the prediction of the statistically significant differences is considerably better than random. This measure depends somewhat on the significance level: For example, τ_{90} is still 0.66 (based on 29 pairs), but τ_{80} = 0.42 (based on 38 pairs). The standard errors of the predictions are 1–3 kJ/mol, which is comparable to the experimental uncertainty (1.7 kJ/mol).

The MM/PBSA predictions show an even larger range, 91 kJ/mol, but the predictions are more centred around the experimental result (the mean signed deviation, MSD, is –4 kJ/mol). From the scatter plot in Figure 3b (squares), it can be seen that ligands L15, L16, and L27 still give too positive affinities. This poor binding arises from the electrostatic, polar solvation terms, and entropy terms. The MADtr (20 kJ/mol) and τ_{95} (0.43, based on 21 pairs) are slightly worse than for the MM/GBSA method, but τ and the correlation is similar. The standard error is slightly larger than for MM/GBSA, 1–4 kJ/mol.

The LIE predictions are on average 20 kJ/mol more positive than the experimental energies and they show a range of 37 kJ/mol. However, ligands L15 and L13 are a prominent outliers (cf. the scatter plot in Figure 3c, squares) and without these ligands, the range is only 13 kJ/mol (6 kJ/mol for the active compounds in Figure 3c, i.e. smaller than the experimental range). MADtr is 4 kJ/mol, i.e. much better than MM/GBSA and MM/PBSA, reflecting the smaller range of the affinities. However, Kendall's τ (0.01) and r^2 (0.02) are still very small. Because of the small range of the calculated affinities, only 8 of the ligand pairs have a significant affinity difference, but for these the ranking is correct, τ_{95} = 1.0 (but τ_{80} = 0.57; based on 23 pairs). The standard deviations are slightly smaller than for MM/GBSA, 1–2 kJ/mol.

The LIE method has one adjustable parameter, α, which is often optimised to improve the accuracy of the predictions, although it has been claimed that a value of 0.18 is universal [60]. In a blind test, it is not possible to optimise the parameter. However, in retrospect, when the experimental affinities have been released, one could do such a fitting. We have optimised α against MADtr (which gave best results in a our previous investigation [20]) and it gave α = 0.39. Using this value, the LIE estimates become 11 kJ/mol more negative on average and τ is improved to 0.3 (τ_{95} does not change). However, MADtr is only improved by 0.3 kJ/mol and there is still no correlation to the experimental results (r^2 = 0.06).

The CLIE method gives a range (33 kJ/mol) and an offset compared to experiments (–21 kJ/mol) that are similar to those of the LIE method. However, there is only a rather weak correlation between the two set of predictions (r^2 = 0.34 for all 34 ligands). L15 and L16 give the weakest affinities, but they are rather close to the other estimates (Figure 3d, squares). The MADtr of the CLIE method is slightly worse than that of LIE, 6 kJ/mol, and τ = –0.04 and r^2

= 0.00 are also poor. τ_{95} = 0.18 (based on 17 pairs) is also worse than for the other methods. The standard deviations are slightly lower than for LIE. Optimising the α parameter against MADtr, gives α = 1.0, i.e. a different value than for LIE. This makes the predictions on average about 41 kJ/mol more negative. *τ* then improves to 0.2, τ_{95} increases to 0.46 (based on 11 pairs; τ_{80} = 0.50 with 24 pairs), and MADtr decreases to 4 kJ/mol, but r^2 is still only 0.03.

The Glide scoring shows a range similar to that of the LIE methods (24 kJ/mol), but this is caused entirely by the pK_a correction. Without ligands L1, L15, and L25, the range is 12 kJ/mol for all ligands and only 5 kJ/mol for the active ligands, i.e. less than the experimental range. On average, the predictions are 10 kJ/mol more negative than the experimental results. Glide scoring is the only method that gives a MADtr comparable with that of the null hypothesis, 3 kJ/mol. However, *τ* = 0.2 and the correlation is still poor, r^2 = 0.02. The Glide scoring is not based on any MD simulations and therefore does not involve any uncertainty owing to different structures. We therefore, calculated τ_{95} based on all 29 pairs that have experimentally significant differences, which gave τ_{95} = 0.52 (τ_{80} = 0.54, based on 43 pairs).

In order to understand the outliers better and to examine how the docking pose and protonation state of the ligands affect the results, we performed additional simulations of ligands L13, L15, L16, and L27. For L13, we neutralised the nitrogen atom in the aliphatic ring, giving it a net charge of +1. As can be seen in Table 1, this resulted in statistical significant differences for all methods except CLIE, and the affinity became more negative than when the ligand had a +2 charge. This ensured that it is no longer an outlier in the LIE scatter plot (Figure 3, triangles). It also came closer to the other ligands with similar affinities for the MM/GBSA and MM/PBSA methods.

For L15, we protonated the nitrogen atom that is only in the five-ring and deprotonated the carboxylate group (thereby avoiding the pK_a correction). Moreover, the molecule was turned around so that the protonated nitrogen pointed towards Asp-189. These modification gave rise to statistically significant differences for all methods and in all cases, the modifications also led to a lower (more negative) free energy. This ensured that this ligand is no longer an outlier for LIE and CLIE, although it is still an outlier for MM/GBSA and MM/PBSA (but it comes closer to the other points).

For L16, three new sets of simulations were performed. In the first set, the nitrogen atom was neutralised, making the molecule neutral. In the second set, the ligand was kept charged but it was turned around such that the hydroxyl group interacted with Asp-189. In the third set of simulations, the ligand was both neutralised and turned around. The neutralisation gave only small changes in the binding affinity 1–4 kJ/mol, expect for MM/PBSA, for which a 20 kJ/mol more negative binding affinity was obtained. The effect of the alternative binding mode varied among the various methods, with differences from 2 kJ/mol for the optimised CLIE to 17 kJ/mol for LIE. Thus, L16 remains an outlier for MM/GBSA and also for MM/PBSA, although in the latter case it moves closer to the other points. On the other hand, the alternative binding mode would become a prominent outlier for LIE, with a too low affinity, whereas it would move closer to the other ligands for CLIE. Neutralising the ligand in the alternative binding mode did not change the results significantly, except for MM/PBSA.

For L27, we tried to protonate the NH_2 group, giving the ligand a net positive charge. This resulted in statistically significant changes for all methods, giving it a more negative affinity. This ensured that L27 is no longer an outlier in the MM/GBSA, MM/PBSA, and optimum CLIE scatter plot, whereas the results for LIE were somewhat deteriorated.

Thus, we can conclude that the outliers for LIE and CLIE can be removed by modifying the protonation of the ligands, whereas for MM/GBSA and MM/PBSA some outliers remain and modifications that improve the results for these methods often give worse predictions with LIE. Moreover, even if the outliers for LIE or CLIE are removed, the quality measures are not significantly improved. On the other hand, these calculations show us what effects can be expected when the protonation state or binding orientation of the ligands are modified.

Another interesting question is how the various methods manage to discriminate between ligands that experimentally are assigned binders or non-binders (although it should be noted that this division is somewhat artificial and arbitrary; the experimental method cannot measure binding affinities less negative than –17 kJ/mol, but the worst binder has an affinity only 0.4 kJ/mol above this limit, i.e. within experimental uncertainty, and the best binder has only 9 kJ/mol stronger affinity). We have computed receiver-operating-characteristic (ROC) curves for all five methods, which are shown in Figure 4. In Table 1, the computed ROC area under the curve (AUC) is summarised. It can be seen that the LIE results never go above the line of no discrimination and therefore actually is worse than a random guess at distinguishing active and inactive ligands (AUC = 0.3). The CLIE method displays a slightly better discrimination power, but it is similar to a random guess with an AUC of 0.51. The AUC values for LIE and CLIE are not changed when the new simulations of L13 and L15 is used. The MM/GBSA and MM/PBSA methods are better and show both an AUC of 0.7. However, all of the simulation-based methods are outperformed by the Glide score, which shows an impressive level-off at a sensitivity of 0.95 already at a false-positive rate of ~0.3. The AUC for the Glide score is 0.8.

The host–guest challenge

The host–guest challenge involved the binding of seven ligands (G1–G7) to one host and the binding of two ligands G8 and G9 to two additional hosts. The ligands are shown in Figure 2.

Host 1 contains four terminal carboxylate groups that are rather close in space and may interact with the ligand. Therefore, we first needed to decide the protonation state of these residues. We took an empirical approach to this problem by estimating MM/GBSA binding affinities for four published guests (shown in Figure S2). Three different protonation states of the host were considered, one fully protonated (giving a neutral host), one fully deprotonated (giving a host with –4 charge), and one with half of the carboxyl groups protonated (carboxyl groups diagonal to each other where protonated, giving a host with –2 charge). The results are summarised in Table S2. The binding affinities obtained with the various protonation states differ by 1–21 kJ/mol. All methods give the correct ranking of the four guests, except for the G23–G27 pair, thereby giving Kendall's τ = 0.67. However, for the neutral host, the difference in the calculated binding affinity between these two guests is within the statistical uncertainty, so that τ_{95} = 1.0 (based on the remaining 5 pairs), whereas for the two deprotonated hosts, the difference is significant at the 95% level (i.e. τ_{95} = 0.67). The correlation coefficient (r^2) is highest for the neutral host (0.70) and lowest for the fully charged host (0.63). On the other hand, the MADtr ranges from 26 kJ/mol for the neutral host to 15 kJ/mol when using the fully deprotonated host. These are poor MADtr values, considering that the null hypothesis gives 9 kJ/mol. We also observed that the host was more prone that open up when it was fully deprotonated, most likely due to electrostatic repulsion. Therefore, we decided to simulate the Sampl3 guests with a neutral host. This would also allow for a consistent treatment with the LIE approach.

The next task was to decide the protonation state of the guests. Aliphatic amino groups (both linear and cyclic) were assumed to be positively charged, whereas the aniline groups of G2 and G7 were assumed to be neutral. Therefore, all guest were modelled with a single positive charge, except G4 that was neutral and G6 that had a double positive charge. The final protonation states are shown in Figure 2. Both enantiomers of G1 were simulated and their calculated binding free energies were averaged.

The predicted affinities for the binding of G1–G7 to host1 are shown in Table 2. Both the MM/GBSA and MM/PBSA predictions are more negative than the experimental affinities (Figure 5). The range is also considerably larger, 145–148 kJ/mol, compared to the experimental range of 27 kJ/mol. Consequently, the MADtr of 40 and 28 kJ/mol for

MM/GBSA and MM/PBSA, respectively, are much larger than for the null hypothesis of 6 kJ/mol. However, the ranking of the guests is fairly good, with τ values of 0.62 and 0.52 for MM/GBSA and MM/PBSA, respectively (and all pairs had significant differences at the 95% level). The correlation is also decent, r^2 = 0.58 for MM/GBSA and 0.80 for MM/PBSA. On the other hand, the slopes of the best correlation lines is 4.7 for both methods, i.e. much larger than unity. The standard errors are slightly lower than for the trypsin ligands, around 1 kJ/mol, except for G6 with its large size and double charge (4 kJ/mol).

After submission, the host–guest simulations were complemented with simulations of the free guest in water, so that LIE and CLIE estimates could also be computed. These are included in Table 2, as well. The range of the LIE results are much smaller than the MM/GBSA and MM/PBSA results, 54 kJ/mol. Therefore, the MADtr is much smaller as well, 11 kJ/mol. However, the ranking of the guests is significantly worse, with a τ of 0.24 ($τ_{95}$ = 0.27). The correlation coefficient is 0.59 and the slope is 1.8.

We also optimised the α parameter with respect to MADtr. This gave α = 0.29, but MADtr was only decreased by 0.1 kJ/mol. On the other hand, τ was improved to 0.33 ($τ_{95}$ = 0.44) and r^2 was also somewhat increased to 0.64.

Interestingly, the range of the CLIE affinities is 90 kJ/mol, i.e. intermediate between the LIE and the MM/GB(PB)SA results. Consequently, the MADtr is also intermediate, 17 kJ/mol. However the $τ_{95}$ of 0.50 is similar to that of the PBSA results, as is the correlation coefficient, 0.74. When the α parameter was optimised with respect to MADtr, it decreased to 0, i.e. turning off this term so that only the electrostatic interactions remain. This improved MADtr by 0.6 kJ/mol, but all the other quality measures were deteriorated. The standard errors of the CLIE and LIE results are similar to the MM/GBSA and MM/PBSA results.

The Glide score shows a much narrower range of the affinities, 7 kJ/mol, i.e. actually narrower than the experimental affinities. Therefore, the MADtr of 5 kJ/mol is also much better than for the other methods and slightly better than for the null hypothesis. However, the correlation is worse than for the other methods, r^2 = 0.3 and τ (= $τ_{95}$) is 0.43.

MM/GBSA and MM/PBSA predictions of the affinities of the guests for host2 and host3 are also more negative than the experimental results. This is especially pronounced for host2, for which the GB results are more than 200 kJ/mol too negative. Both the MM/GBSA and MM/PBSA methods predict correctly that G8 binds more weakly than G9 to host2, whereas the reverse is true for host3. However, the predicted energy differences are 3–7 times too large. Moreover, none of the methods predicts that host2 binds the two guests better than does host3.

The LIE methods predict that G8 binds better than G9 to both host2 and host3, contrary to experiments for host2. It is also notable that the LIE predictions become positive for host3, so that LIE also is unable to predict that host2 binds the guests weaker than host3. The CLIE method predicts the correct binding order of G8 and G9 in the two hosts, but it is also unable to predict the ranking between the two hosts.

Finally, the Glide scoring predicts that G8 binds better than G9 in both hosts, contrary to experiments for host2. However, the Glide scoring correctly predicts that host3 binds its guests stronger than host2.

Conclusions

In this article, we present binding affinities estimated with the MM/GBSA, MM/PBSA, LIE, CLIE, and Glide scoring methods for the SAMPL3 blind test, involving the binding of 34 small ligands to trypsin and nine guests to three small host molecules. For the active trypsin ligands, no method provides any reliable results, giving correlation coefficients below 0.1 and Kendall's τ values below 0.3. In particular, no method could point out the three best binders. Uncertainties in the protonation state of the protein and the ligands affect the results and some outliers can be avoided by assuming different protonation states, especially for LIE and CLIE, but the quality measures were not improved significantly. However, the most likely reason for this poor performance is the small range of experimental affinities (only 9 kJ/mol). In fact, only 29 pairs (of 135 possible) of ligands have an experimental difference in affinity that is significant at the 95% level. We have introduced a new metric, τ_{95} , that considers only pairs that have statistically significant differences both in the calculations and in the experiments. τ_{95} for the various methods ranges from 0.2 to 1.0, showing that most of the methods can rank the ligands that have statistically significant differences better than random. Moreover, all methods, except LIE were successful in distinguishing the active and inactive ligands, giving AUCs of 0.7 for MM/PB(GB)SA and AUC = 0.8 for Glide score.

For the host–guest binding, MM/GBSA and MM/PBSA gave reasonable results, e.g. correlation coefficients of 0.6–0.8 and τ values of 0.5–0.6. LIE and CLIE gave almost as good results, whereas Glide scoring gave slightly worse results, showing that the performance of the various methods depends quite strongly on the test system considered. The reason for this might be that Glide was optimised for protein-ligand complexes.

It is notable that the MM/PB(GB)SA methods give a too large range of estimated affinities. This indicates that these methods could gain somewhat from scaling down the electrostatic interactions by a dielectric constant, although this is not the case for some other proteins [61]. LIE gives results of a more reasonable range. The results of this method can also be somewhat improved by optimising the α parameter, but the improvement is modest and the optimum values of α varies widely among the different systems.

Acknowledgements

This investigation has been supported by grants from the Swedish research council (project 2010-5025), the Danish council for independent research, and from the FLÄK research school in pharmaceutical science at Lund University. It has also been supported by computer resources of Lunarc at Lund University, HPC2N at Umeå University, and C3SE at Chalmers Technical University.

References

1 Gohlke H, Klebe G (2002) Angew Chem Int 41:2644-2676.

2 Moghaddam, S, Inoue, Y, Gilson, M K (2009) J Am Chem Soc., 131:4012-4021.

3 Lee FS, Chu ZT, Bolger MB, Warshel A (1992) Protein Eng 5:215-228.

4 Singh N, Warshel A (2010) Proteins 78:1705-1723.

5 Åqvist J, Medina C, Samuelsson J-E (1994) Protein Eng 7:385-391

6 Hansson T, Marelius J, Åqvist J (1998) J Comput-Aided Mol Design 12:27-35

7 Brandsdal B O, Östberg FM, Almlöf M, Feierberg I, Luzhkov VB, Åqvist J (2003) Adv Prot Chem 66:123-158

8 Srinivasan, J, Cheatham III, T E, Cieplak, P, Kollman, P A, Case, D A (1998) J Am Chem Soc 37:9401-9409

9 Kollman, P A, Massova, I, Reyes, I, Kuhn, B, Huo, S, Chong, L, Lee, M, Lee, T, Duan, Y, Wang, W, Donini, O, Cieplak, P, Srinivasan, J, Case, D A, Cheatham III, T E (2000) Acc Chem Res 33:889-897

10 Genheden S, Ryde U (2010) J Comput Chem 31:837-846

11 Genheden S, Ryde U (2011) J Comput Chem 32:187-195

12 Kongsted J, Ryde U (2009) J Comput Aided Mol Design 23:63-71

13 Weis A, Katebzadeh K, Söderhjelm P, Nilsson I, Ryde U (2006) J Med Chem 49:6596-6606

14 Genheden S, Söderhjelm P, Ryde U (2010) Int J Quant Chem, in press (doi 10.1002/qua.22967).

15 Kongsted J, Söderhjelm P, Ryde U (2009) J Comp-Aided Mol Design 23:395-409

16 Genheden S, Luchko T, Gusarov S, Kovalenko A, Ryde U (2010) J Phys Chem B, 114:8505-8516

17 Genheden S, Kongsted J, Söderhjelm P, Ryde U (2010) J Chem Theory Comput 6:3558-3568

18 Genheden S, Mikulskis P, Hu L, Kongsted J, Söderhjelm P, Ryde U (2011) J Am Chem Soc 133:13081-13092

19 Söderhjelm P, Kongsted J, Ryde U (2010) J Chem Theory Comput, 6:1726-1737

20 Genheden S, Ryde U (2011) J Chem Theory Comput, in press (doi 10.1021/ct200163)

21 Genheden, S (2011) J Comp-Aided Mol Design, submitted

22 Lawrenz M, Baron R, McCammon J A (2009) J Chem Theory Comput 5:1106-1116

23 Barril X, Gelpi J L, Lopez J M, Orozco M, Luque F J (2001) Theor Chem Acc 106:2-9

24 Hou T, Guo S, Xu, X (2002) J Phys Chem B 106:5527-5535

25 Salvalagli M, Zamolo L, Busini V, Moscatelli D, Cavallotti C (2009) J Chrom A 1216:8678-8686

26 Guthrie JP (2009) J Phys Chem B 113:4501-4507

27 Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) J Comput Aided Mol Des 24:259-279.

28 SAMPL3, overview.

29 Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) J Med Chem 49:6177-6196

30 Shelley, JC, Cholleti A, Frye LL, Greenwood, JR, Timlin MR, Uchiyama M, (2007) J Comput Aided Mol Des, 21:681–691.

31 Becke AD (1988) Phys Rev A 38:3098-3100

32 Lee C T, Yang W T, Parr, R G (1988) Phys Rev B, 37:785-789

33 Becke AD (1993) J Chem Phys 98:5648-5652

34 Jaguar, version 7.7, Schrödinger, LLC, New York, 2010

35 Klicic, J J, Friesner RA, Liu S-Y, Guida W C (2002) J Phys Chem A 106:1327

36 Maestro, version 9.1, Schrödinger, LLC, New York, 2010

37 Glide, version 5.6, Schrödinger, LCC, New York, 2010

38 Leiros HK, McSweeney, SM, Smalås, AO. (2001) Acta Crystallogr D Biol Crystallogr. 57:488-97.

39 Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Proteins: 65:712-725.

40 Wang, J M, Wolf, R M, Caldwell, K W, Kollman, P A, Case, D A (2004) J Comput Chem., 25:1157-1174

41 Bayly, C I, Cieplak, P, Cornell, W D, Kollman, P A (1993) J Phys Chem 97:10269-10280

42 Besler, B H, Merz, KM Kollman, P A (1990) J Comput Chem 11:431-439

43 Jorgensen, W L, Chandrasekhar, J, Madura, J D, Impley, R W, Klein, M L (1983) J Chem Phys 79:926-935

44 Jakalian, A, Jack, D B, Bayly, C I (2002) J Comput Chem 23:1623-1641

45 Case, D A, Darden, T A, Cheatham III, T E, Simmerling, C L, Wang, J, Duke, R E, Luo, R, Crowley, M, Walker, R C, Zhang, W, Merz, K M, Wang, B, Hayik, S, Roitberg, A, Seabra, G, Kolossvary, I, Wong, K F, Paesani, F, Vanicek, J, Wu, X, Brozell, S R, Steinbrecher, T, Gohlke, H, Yang, L, Tan, C, Mongan, J, Hornak V, Cui, G, Mathews, D H, Seetin, M G, Sagui, C, Babin, V, Kollman, P A Amber 10, University of California, San Francisco, 2008.

46 Wu, X, Brooks, B.R (2003) Chem Phys Lett 381:512-518

47 Berendsen, H J C, Postma, J P M, Van Gunsteren, W F, Dinola, A, Haak, J R (1984) J Chem Phys 81:3684–3690

48 Darden, T.,;York, D, Pedersen, L (1993) J Chem Phys 98:10089-10092

49 Ryckaert, J P, Ciccotti, G, Berendsen, H J C (1977) J Comput Phys 23.327-341

50 Onufriev, A, Bashford, D, Case, D A (2004) Proteins 55:383-394

51 Carlsson J, Andér M, Nervall M, Åqvist J (2006) J Phys Chem B 110:12034-12041

52 Abdi, H. (2007) In *Encyclopedia of Measurement and Statistics.* Salkind, N. Ed., Sage: Thousand Oaks, CA,

53 Mackman R L, Katz B A, Breitenbucher J G, Hui H C, Verner E, Luong C, Liu L, Sprengeler P A (2001) J Med Chem 44:3856– 3871

54 Matter H, Defossa E, Heinelt U, Blohm P-M, Schneider D, Müller A, Hreok Si, Schreuder H, Liesum A, Brachvogel V, Lonze P, Walser A, Al-Obeidi F, Wildgoose P (2002) J Med Chem 45:2749– 2769

55 Jiao D, Zhang J, Duke R E, Li G, Schnieders M J, Ren P (2009) J Comput Chem 30:1701-1711

56 McGrath M E, Sprengeler P A, Hirschbein B, Somoza J R, Lehoux I, Janc J W, Gjerstad E, Graupe M, Estiarte A, Venkataramani C, Liu Y, Yee R, Ho J D, Green M J, Lee C-S, Liu L, Tai V, Spencer J, Sperandio D, Katz B A (2006) Biochem 45:5964-5973

57 Brandsdal BO, Smalås AO, Åqvist J (2006) Proteins 64:740-748

58 Weast RC Eds. (1974) *CRC Handbook of Chemistry and Physics 54th Ed.*

59 Bode W, P Schwager (1975) *J Mol Biol* 98:683-717

60 Almlöf M, Brandsdal B O, Åqvist, J (2004) J Comput Chem 25:1242-1254

61 Genheden S, Ryde U (2011) Comparison of end-point continuum-solvent methods for the calculation of protein–ligand binding free energies. *Proteins,* submitted.

Table 1. Predicted binding affinities (kJ/mol) of the 34 trypsin ligands with the various methods. Results marked with footnotes a, b, and d-j were obtained after the experimental results were presented.

Ligand	MM/GBSA	MM/PBSA	LIE	LIE[a]	CLIE	CLIE[b]	Glide	Expt[c]
L1	-33.1 ±2.1	20.0 ±3.1	0.6 ±0.9	-9.5 ±0.8	13.0 ±1.0	-26.4 ±1.2	-12.4	NB
L2	-56.4 ±1.4	-33.3 ±2.1	-3.8 ±1.4	-14.6 ±1.4	0.3 ±0.7	-41.7 ±1.0	-32.9	-19.4
L3	-7.3 ±1.5	16.2 ±2.0	-0.8 ±0.5	-6.5 ±0.5	5.7 ±0.5	-16.6 ±1.3	-26.0	NB
L4	-55.9 ±1.1	-27.9 ±1.6	-4.4 ±1.1	-13.7 ±1.0	-6.2 ±0.6	-42.3 ±0.9	-31.0	-17.5
L5	-28.7 ±1.4	6.6 ±2.7	-7.2 ±0.8	-16.9 ±0.7	4.0 ±0.7	-33.8 ±1.0	-23.2	NB
L6	-54.0 ±1.9	-21.5 ±3.4	-5.2 ±1.4	-14.8 ±1.4	-3.0 ±1.2	-40.5 ±1.2	-32.5	-25.7
L7	-29.7 ±1.6	-3.7 ±2.8	-1.8 ±1.0	-14.0 ±1.0	3.1 ±0.8	-44.5 ±1.2	-28.9	-21.8
L8	-58.3 ±1.7	-30.8 ±2.6	-0.8 ±1.2	-11.2 ±1.2	-1.2 ±0.8	-41.9 ±1.3	-30.8	-18.3
L9	-32.2 ±2.6	-10.5 ±3.1	-8.7 ±0.8	-18.8 ±0.8	3.8 ±0.9	-35.9 ±2.1	-23.8	NB
L10	-57.4 ±1.7	-28.7 ±2.2	-5.4 ±1.0	-14.9 ±1.0	-5.6 ±0.9	-42.7 ±1.4	-31.3	-19.5
L11	-53.8 ±1.9	-27.1 ±2.2	-3.8 ±1.2	-13.3 ±1.2	-3.6 ±0.7	-40.4 ±1.2	-32.6	-19.0
L11[d]	-55.1 ±1.7	-28.6 ±2.8	-5.3 ±1.3	-14.7 ±1.3	-4.0 ±0.9	-42.6 ±1.2		-19.0
L12	-62.9 ±1.5	-36.1 ±2.0	-3.7 ±1.3	-14.5 ±1.3	-1.7 ±1.0	-43.8 ±1.3	-32.6	-26.3
L13	-47.2 ±2.8	-11.6 ±2.5	12.8 ±2.1	2.9 ±2.1	-0.5 ±1.0	-39.2 ±1.8	-30.2	-19.2
L13[e]	-62.3 ±2.5	-32.9 ±2.8	-2.2 ±1.7	-13.9 ±1.1	1.3 ±0.9	-43.6 ±1.5		-19.2
L14	-41.0 ±3.0	5.8 ±2.9	-1.7 ±1.2	-12.5 ±1.2	-4.4 ±1.6	-46.4 ±2.3	-22.6	NB
L15	0.2 ±1.8	48.8 ±2.1	25.8 ±1.2	-2.3 ±1.1	21.3 ±1.0	-26.6 ±1.3	-12.5	-20.2
L15[f]	-26.1 ±1.8	7.7 ±2.8	-5.1 ±0.9	-18.5 ±0.8	6.9 ±0.7	-45.2 ±1.2		-20.2
L16	-8.2 ±1.3	39.6 ±2.2	0.1 ±1.3	-13.4 ±1.3	14.3 ±0.8	-38.5 ±1.1	-33.5	-21.2
L16[g]	-6.9 ±3.1	19.8 ±3.7	-1.8 ±1.1	-14.0 ±1.0	10.4 ±1.1	-36.9 ±2.1		-21.2
L16[h]	-9.3 ±2.1	31.9 ±2.8	-17.3 ±1.6	-29.0 ±1.5	6.1 ±1.3	-39.4 ±1.5		-21.2
L16[i]	-6.5 ±2.1	39.1 ±2.7	-16.6 ±1.6	-28.2 ±1.6	8.5 ±1.4	-36.9 ±1.6		-21.2
L17	-61.7 ±1.7	-35.8 ±2.2	-4.1 ±1.3	-13.8 ±1.3	-6.9 ±0.7	-44.7 ±1.3	-33.3	-23.6
L18	-43.1 ±2.3	-2.4 ±3.0	-9.6 ±1.0	-20.7 ±1.0	-4.0 ±1.1	-47.4 ±1.2	-26.5	NB
L19	-56.8 ±2.2	-20.7 ±3.6	-7.3 ±1.2	-17.4 ±1.3	-7.1 ±1.1	-46.4 ±1.4	-21.3	NB
L20	-52.3 ±1.4	-25.4 ±1.7	-4.7 ±0.9	-13.7 ±0.9	-2.3 ±0.8	-37.5 ±1.0	-32.6	NB
L21	-57.2 ±2.6	-16.8 ±2.2	-2.3 ±0.6	-15.0 ±0.6	-3.7 ±0.7	-53.5 ±1.8	-27.4	NB
L22	-50.8 ±1.6	-30.0 ±3.1	-6.3 ±1.3	-14.9 ±1.3	-3.5 ±0.7	-36.8 ±1.1	-32.3	NB
L23	-67.9 ±1.3	-31.0 ±2.2	-3.3 ±1.2	-15.3 ±1.2	-11.9 ±0.7	-59.0 ±1.0	-28.4	NB
L24	-50.2 ±2.8	-2.2 ±2.7	-10.8 ±0.7	-22.4 ±0.8	-3.8 ±1.0	-49.0 ±1.8	-22.6	NB
L25	-26.9 ±2.8	24.1 ±3.2	3.1 ±1.0	-8.1 ±0.9	16.4 ±1.3	-27.2 ±1.5	-9.1	NB
L26	-48.5 ±2.1	-21.4 ±2.6	-0.8 ±1.1	-11.3 ±1.0	0.2 ±0.9	-40.9 ±1.5	-33.2	-20.7
L27	-16.2 ±1.6	14.1 ±2.1	-2.8 ±0.6	-11.6 ±0.6	8.2 ±0.6	-26.3 ±1.1	-29.3	-19.0
L27[j]	-58.5 ±1.9	-13.4 ±2.7	-8.3 ±1.3	-19.1 ±1.3	-1.7 ±0.8	-44.1 ±1.1		-19.0
L28	-48.2 ±1.0	-17.2 ±1.4	-9.0 ±1.5	-18.8 ±1.5	-2.6 ±0.5	-41.2 ±0.8	-32.2	NB
L29	-67.9 ±1.8	-42.5 ±2.5	-0.6 ±1.3	-12.2 ±1.2	-6.3 ±0.7	-51.6 ±1.1	-31.9	-20.3
L30	-63.5 ±1.7	-29.9 ±2.3	-6.3 ±1.0	-17.0 ±1.0	-2.8 ±0.8	-45.0 ±1.0	-30.3	-19.9
L31	-52.0 ±1.1	-23.3 ±1.6	-8.0 ±1.2	-17.9 ±1.2	-3.3 ±0.5	-42.1 ±0.7	-32.3	NB
L32	-57.4 ±1.7	-31.6 ±2.6	-3.8 ±1.2	-15.4 ±1.2	5.1 ±0.8	-40.0 ±1.2	-29.4	NB
L33	-57.8 ±1.8	-30.4 ±2.1	-3.7 ±1.0	-12.9 ±1.0	-6.7 ±0.9	-42.5 ±1.2	-31.0	-19.5
L34	-21.5 ±1.7	2.1 ±3.1	-1.9 ±0.7	-10.2 ±0.6	7.0 ±0.6	-25.7 ±1.3	-24.1	NB
range	68.1	91.3	36.7	25.3	33.2	42.4	24.4	9.2
AUC[k]	0.65 ±0.02	0.68 ±0.02	0.31 ±0.03	0.29 ±0.03	0.51 ±0.02	0.54 ±0.03	0.79 ±0.05	
MADtr	15.7 ±0.5	19.7 ±0.6	3.8 ±0.4	3.5 ±0.4	6.1 ±0.4	4.4 ±0.4	3.2 ±0.5	
τ	0.10 ±0.13	0.05 ±0.13	0.01 ±0.13	0.27 ±0.14	-0.04 ±0.13	0.21 ±0.13	0.26 ±0.15	
τ_{95}[l]	0.70 ±0.05 (20)	0.43 ±0.06 (21)	1.00 ±0.08 (8)	1.00 ±0.10 (8)	0.18 ±0.06 (17)	0.46 ±0.07 (11)	0.52 ±0.03 (29)	
r^2	0.01 ±0.04	0.01 ±0.04	0.02 ±0.05	0.06 ±0.06	0.00 ±0.03	0.03 ±0.05	0.02 ±0.05	
slope	0.83 ±1.09	0.83 ±1.40	0.39 ±0.28	0.48 ±0.27	0.11 ±0.41	0.46 ±0.34	0.31 ±0.29	

[a] Using an optimised α = 0.38.

[b] Using an optimised α = 1.0.

[c] Experimental binding affinity; NB = no binding detected.

[d] Results obtained with the binding mode observed in the crystal structure.

[e] Results obtained with a neutral aliphatic ring

[f] Results obtained with a binding mode where the nitrogen-containing ring points towards Asp-189 and the carboxylate group towards the solution. In addition, the nitrogen atom was protonated and the carboxylate group was deprotonated

[g] Results obtained with a neutral ligand

[h] Results obtained with a binding mode where the hydroxyl group points towards Asp-189

[i] Results obtained with a neutral ligand and a binding mode where the hydroxyl group points towards Asp-189

[j] Results obtained with a charged ligand

[k] Area under the ROC curve (see Figure 2), i.e. the probability that the active compounds have been ranked higher than the inactive compounds.

[l] The Kendall's τ rank correlation coefficient, including only pairs with a statistically significant difference at the 95% confidence level. The number of such pairs is given in brackets.

Table 2. Predicted binding affinities of the Sampl3 host–guest challenge (kJ/mol). The LIE and CLIE results were obtained after the experimental results were presented.

	MM/GBSA	MM/PBSA	LIE	LIE[a]	CLIE	CLIE[b]	Glide	Exp
Host1								
G1	-117.1 ±1.0	-99.2 ±1.1	-12.1 ±0.8	-21.6 ±0.7	-27.5 ±0.9	-12.0 ±0.9	-20.8	-24.4 ±0.1
G2	-128.2 ±0.4	-111.0 ±0.8	0.7 ±1.5	-10.1 ±1.5	-27.3 ±1.3	-9.6 ±1.3	-24.9	-29.7 ±0.2
G3	-194.4 ±0.7	-141.3 ±0.7	3.2 ±2.5	-8.6 ±2.6	-44.4 ±0.8	-25.1 ±0.8	-21.6	-28.5 ±0.2
G4	-92.8 ±0.5	-58.4 ±0.8	-0.2 ±0.6	-10.0 ±0.5	-6.2 ±0.4	9.8 ±0.4	-24.3	-17.4 ±0.2
G5	-183.4 ±1.7	-125.3 ±1.5	1.4 ±2.0	-7.8 ±2.0	-51.6 ±0.0	-36.7 ±0.0	-21.0	-25.4 ±0.1
G6	-240.5 ±3.6	-203.8 ±3.8	-51.3 ±3.6	-64.2 ±3.7	-96.3 ±4.4	-75.2 ±4.3	-28.2	-44.9 ±0.3
G7	-133.6 ±0.6	-104.9 ±0.8	-1.7 ±1.7	-14.0 ±1.7	-31.2 ±0.8	-11.1 ±0.8	-21.4	-32.8 ±0.2
MADtr	39.9 ±0.7	27.7 ±0.7	10.9 ±0.8	10.8 ±0.8	16.8 ±0.8	16.2 ±0.7	5.0 ±0.2	5.8
τ	0.62 ±0.00	0.52 ±0.00	0.24 ±0.12	0.33 ±0.11	0.43 ±0.05	0.33 ±0.06	0.43 ±0.11	
τ_{95} [c]	0.62 (21) ±0.00	0.52 (21 ±0.00	0.27 (11) ±0.03	0.44 (18) ±0.03	0.50 (20) ±0.01	0.44 (18) ±0.04	0.43 (21) ±0.02	
r^2	0.58 ±0.02	0.80 ±0.01	0.59 ±0.03	0.64 ±0.03	0.74 ±0.02	0.69 ±0.02	0.33 ±0.06	
slope	4.65 ±0.15	4.69 ±0.15	1.77 ±0.14	1.91 ±0.14	2.88 ±0.17	2.65 ±0.16	0.19 ±0.02	
Host2								
G8	-235.3 ±0.3	-92.6 ±0.3	-25.4 ±1.6		-68.9 ±0.4		-14.6	-25.6 ±0.5
G9	-267.7 ±0.4	-111.3 ±0.4	-16.6 ±1.8		-84.4 ±0.4		-4.3	-31.1 ±0.5
Host3								
G8	-188.5 ±0.8	-93.9 ±0.4	6.7 ±2.3		-35.0 ±0.6		-18.1	-40.2 ±0.3
G9	-181.8 ±1.4	-75.6 ±0.7	10.8 ±1.9		-25.7 ±1.0		-15.8	-37.6 ±0.3

[a] Using an optimised α =0.29.

[b] Using an optimised α = 0.0.

[c] The Kendall's τ rank correlation coefficient, including only pairs with a statistically significant difference at the 95% confidence level. The number of such pairs is given in brackets.

Figure 1. Trypsin ligands and the protonation state used in the calculations. The black circle shows which part of the molecule was directed towards Asp-189.

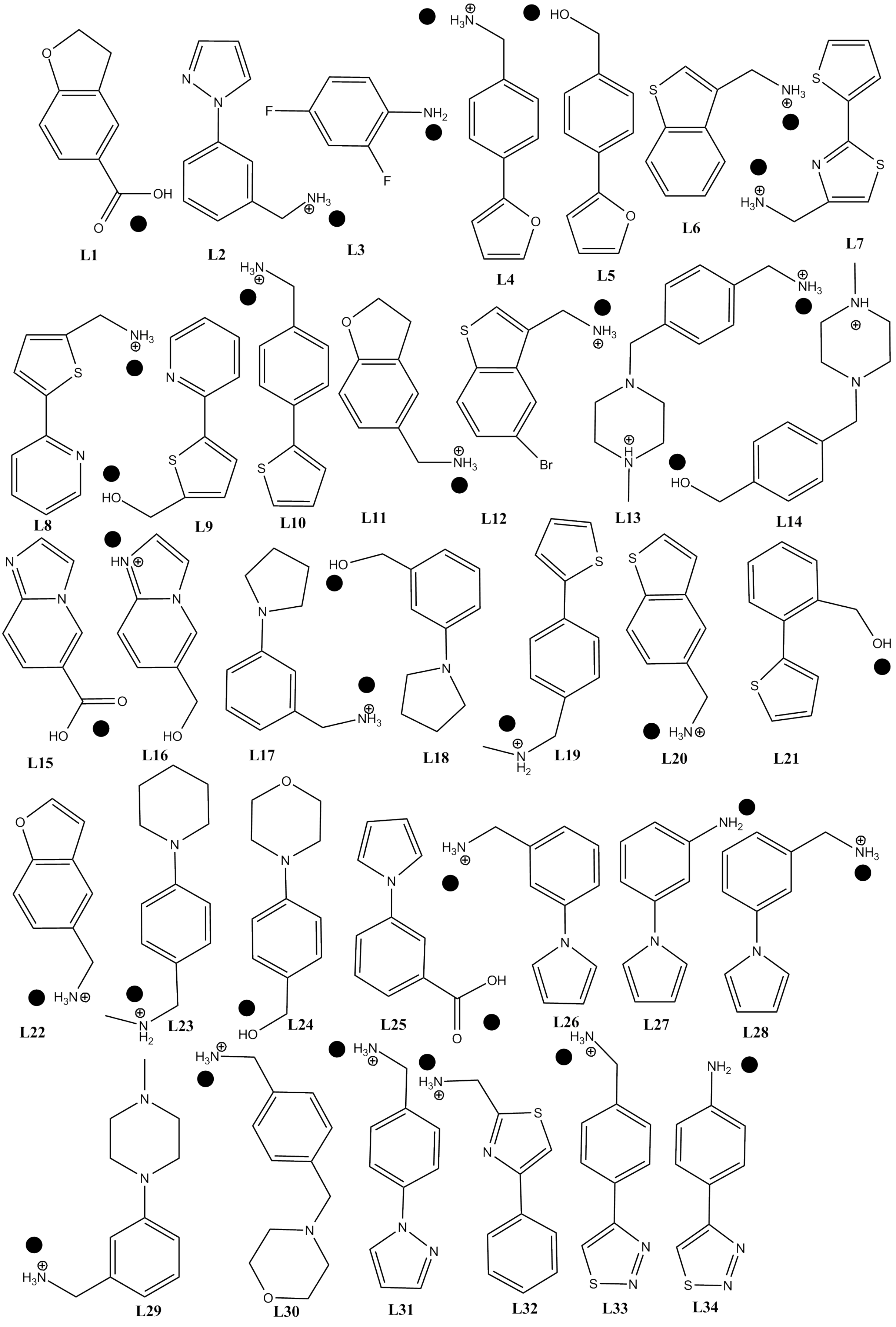


Figure 2. The nine guest molecules in the Sampl3 challenge.

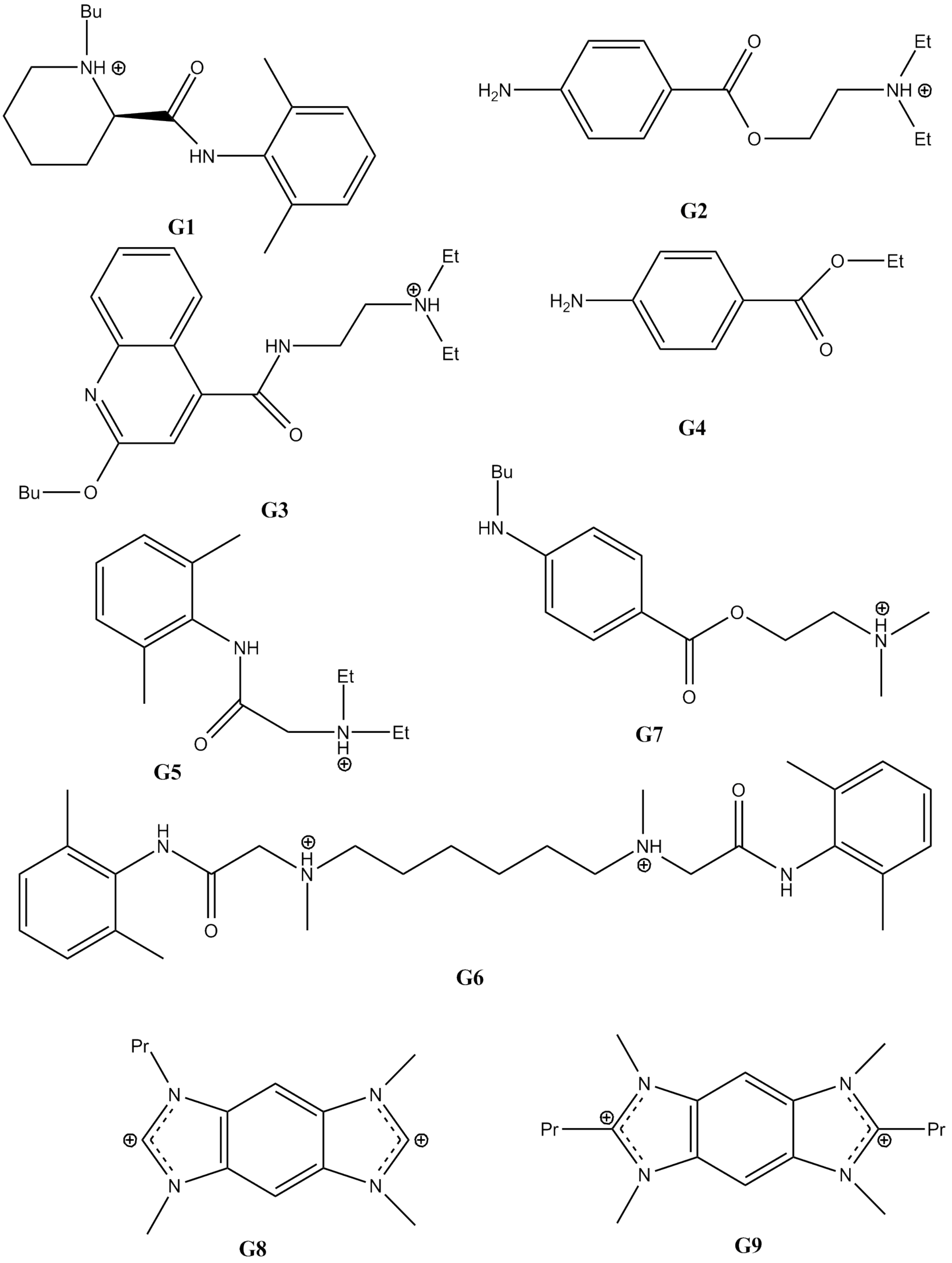


Figure 3. Scatter plots of the calculated and experimental binding affinities for trypsin challenge, using the a) MM/GBSA, b) MM/PBSA, c) LIE, d) CLIE, and e) Glide scoring methods. Original data are shown as squares or diamonds in case of LIE and CLIE with optimised values of α, and recalculated data are shown as triangles. The original affinities for the ligands that were recalculated are shown in a different colour than the other ligands (yellow or light blue).

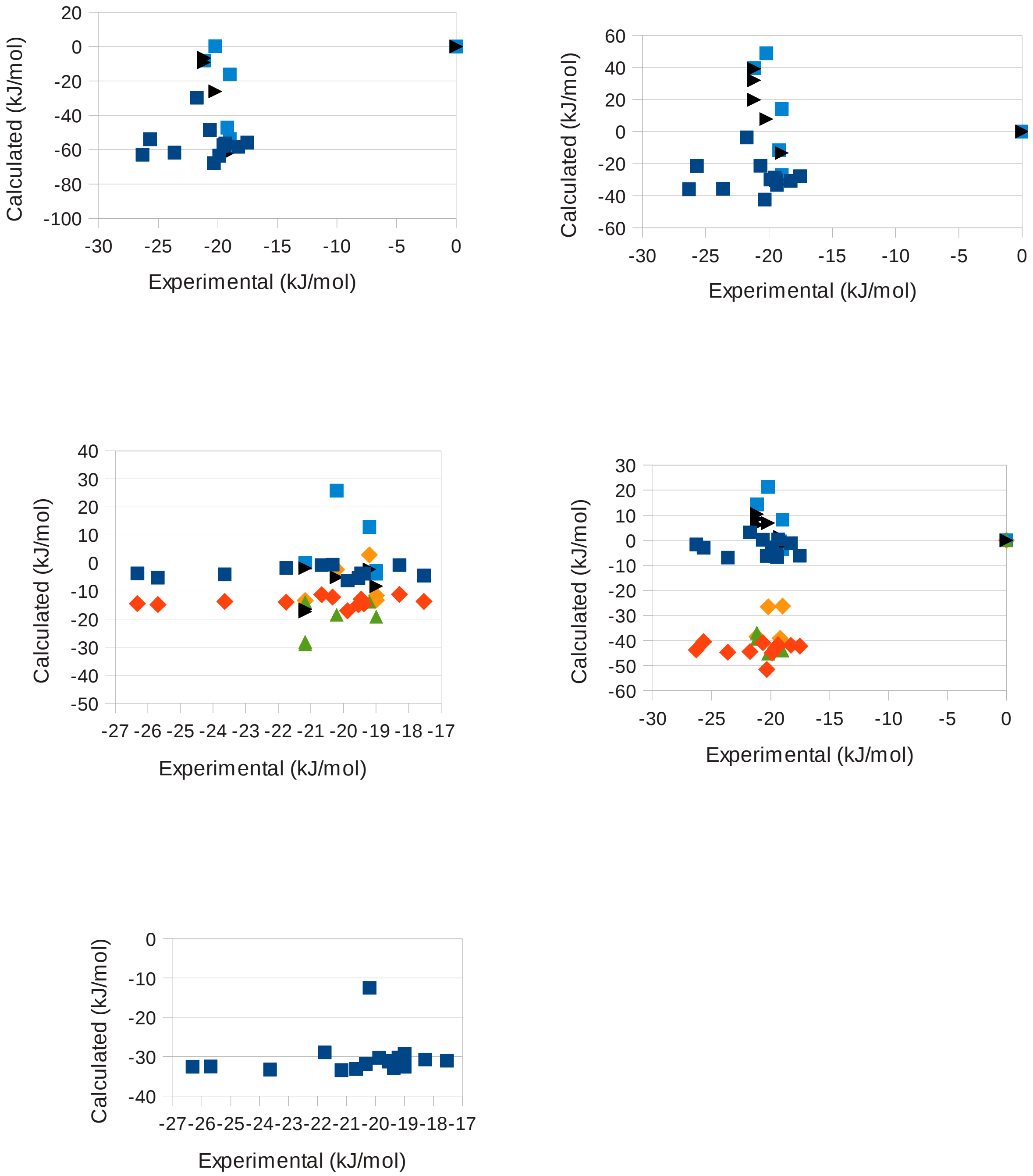


Figure 4. ROC curves for the various methods in the trypsin challenge.

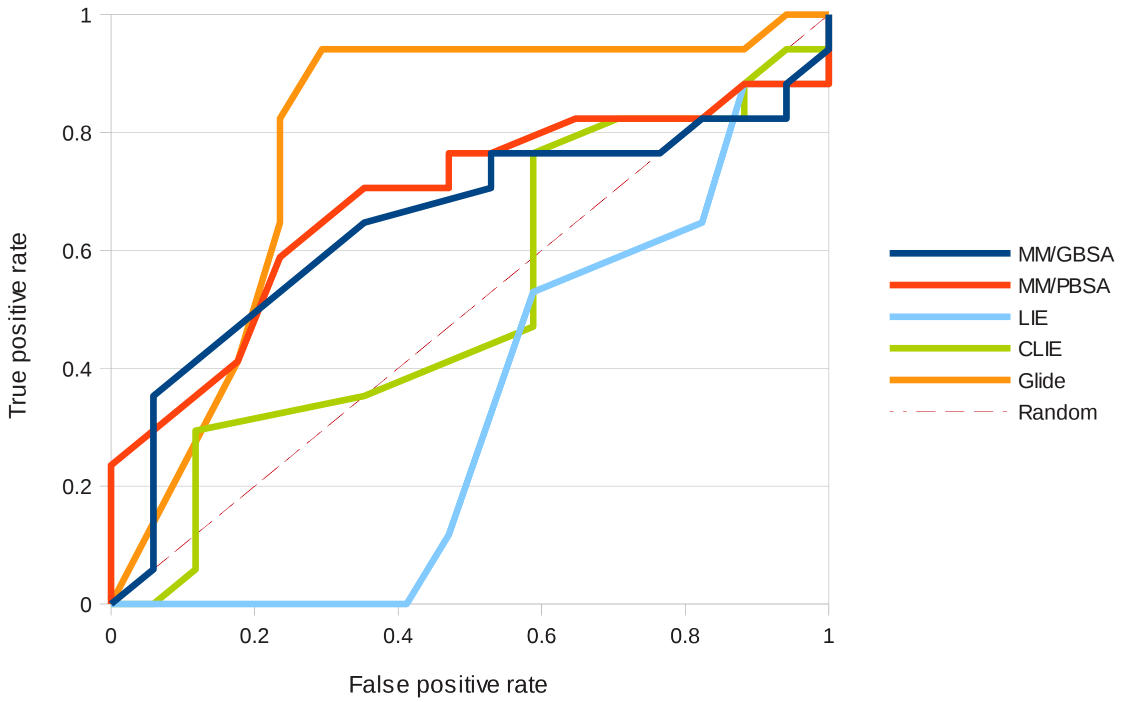
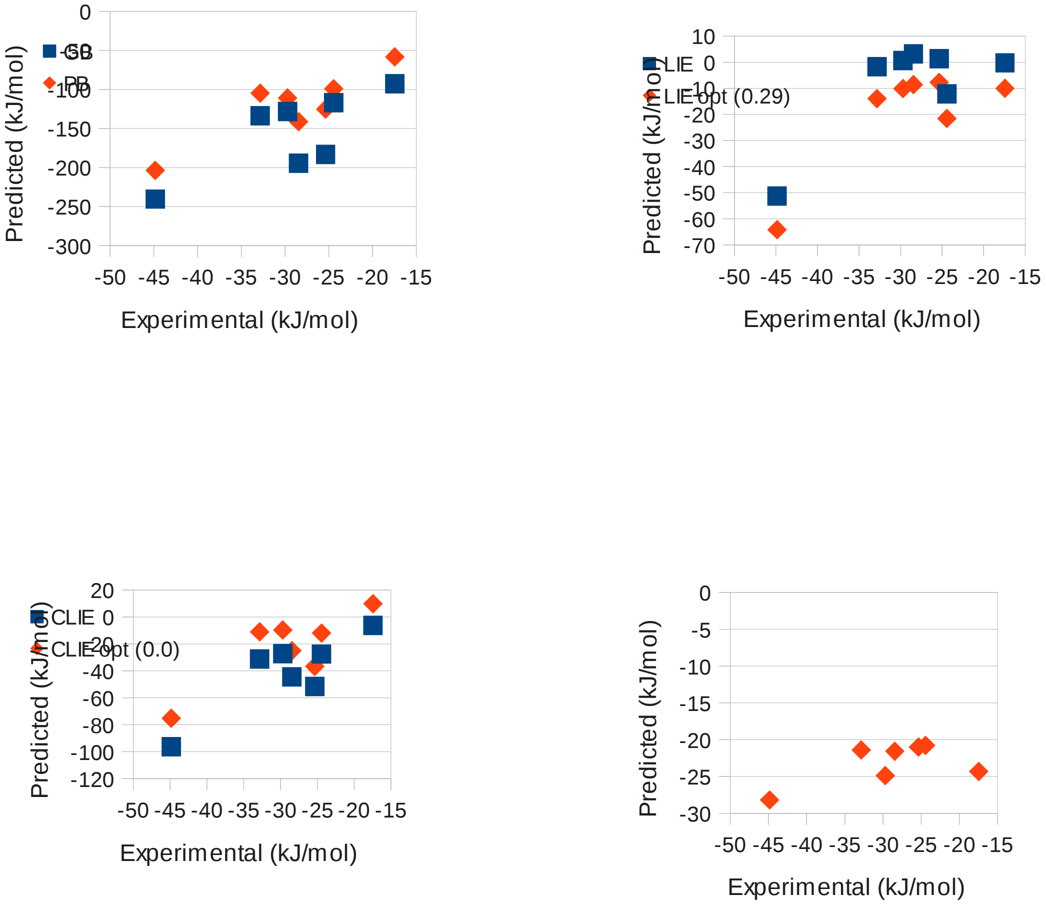


Figure 5. Scatter plots for the host1 predictions with the a) MM/GBSA and MM/PBSA, b) LIE, c) CLIE, and d) Glide scoring methods.


GB
PB
Predicted (kJ/mol)
Experimental (kJ/mol)
LIE
LIE-opt (0.29)
Predicted (kJ/mol)
Experimental (kJ/mol)
CLIE
CLIE-opt (0.0)
Predicted (kJ/mol)
Experimental (kJ/mol)
Predicted (kJ/mol)
Experimental (kJ/mol)