LUND UNIVERSITY

The Normal-Mode Entropy in the MM/GBSA Method: Effect of System Truncation, Buffer Region, and Dielectric Constant

Genheden, Samuel; Kuhn, Oliver; Mikulskis, Paulius; Hoffmann, Daniel; Ryde, Ulf

The normal-mode entropy in the MM/GBSA method: Effect of system truncation, buffer region, and dielectric constant

Samuel Genheden*[1], Oliver Kuhn[2], Paulius Mikulskis[1], Daniel Hoffmann[2], Ulf Ryde[1]

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

[2] Department of Bioinformatics, Center for Medical Biotechnology, University of Duisburg-Essen, Universitätsstraße 1-5, 45117 Essen, Germany

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

2013-01-29

We have performed a systematic study of the entropy term in the MM/GBSA (molecular mechanics combined with generalised Born and surface-area solvation) approach to calculate ligand-binding affinities. The entropies are calculated by a normal-mode analysis of harmonic frequencies from minimised snapshots of molecular dynamics simulations. For computational reasons, these calculations have normally been performed on truncated systems. We have studied the binding of eight inhibitors of blood clotting factor Xa, nine ligands of ferritin, and two ligands of HIV-1 protease and show that removing protein residues with distances larger than 8–16 Å to the ligand and including a 4 Å shell of fixed protein residues and water molecules, change the absolute entropies by 1–5 kJ/mol on average. However, the change is systematic, so relative entropies for different ligands change by only 0.7–1.6 kJ/mol on average. Consequently, entropies from truncated systems give relative binding affinities that are identical to those obtained for the whole protein within statistical uncertainty (1–2 kJ/mol). We have also tested to use a distance-dependent dielectric constant in the minimisation and frequency calculation ($\varepsilon = 4r$), but it typically gives slightly different entropies and poorer binding affinities. Therefore, we recommend entropies calculated with the smallest truncation radius (8 Å) and $\varepsilon = 1$. Such an approach also gives an improved precision for the calculated binding free energies.

Introduction

MM/GBSA is an approximate method to estimate the absolute binding free energy, ΔG_{bind}, of a ligand L, to a biomacromolecule, e.g. a protein, P, forming a complex PL. It estimates ΔG_{bind} as the difference in free energy between PL, P, and L, viz. $\Delta G_{bind} = G(PL) - G(P) - G(L)$. Each of these three free energies are estimated from the following sum[1,2]

$$G = \left\langle E_{int} + E_{vdW} + E_{ele} + G_{solv} + G_{np} - TS_{MM} \right\rangle \tag{1}$$

where the brackets indicate an average over snapshots from a molecular dynamics (MD) simulation. The first three terms on the right-hand side are the molecular mechanics (MM) internal (i.e. bonds, angles, and dihedral), van der Waals, and electrostatic energies, G_{solv} and G_{np} are the polar and non-polar solvation free energies, and the last term is the absolute temperature multiplied by an entropy estimate. The latter estimate is usually taken as a combination of translational, rotational, and vibrational terms. All the terms are calculated on a system where water molecules from the simulation have been stripped off. Moreover, typically only the complex (PL) is simulated and the free energies of P and L are obtained from the same simulation by simply deleting the coordinates of the other species.[3,4] Thereby, the E_{int} term cancels exactly and the precision is improved by a factor of ~5.[5,6]

The most common method to estimate the vibrational entropy in the MM/GBSA method is to use frequencies from a normal-mode analysis (NMA),[7] performed at the MM level. In a few studies, alternative methods have been tested, e.g. counting the number of rotable bonds in the ligand, dihedral-distribution histogramming, or quasi-harmonic analysis.[8,9,10,11,12] The latter two estimates seem to be very hard to converge to any useful precision, whereas NMA estimates of entropies converge much better.[10,11,12] In practice, the NMA calculations are performed by first stripping off all water molecules. Typically, the protein is also truncated by removing all residues more than ~8 Å from the ligand,[2] because the frequency calculations are very demanding in terms of computer time and memory. Then, each truncated snapshot is minimised and harmonic frequencies are calculated. In order to account for solvent effects in an approximate manner (i.e. to substitute for the deleted explicit water molecules), these calculation are often performed with a distance-dependent dielectric constant, viz. $\varepsilon = 4r$.

However, it has been shown that such an approach gives large fluctuations in the calculated entropies and therefore a large statistical uncertainty, which typically limits the overall precision of the MM/GBSA method.[13,14] The reason for this is probably that the minimisation of the truncated and unconstrained system with the approximate $\varepsilon = 4r$ may give rise to large and varying changes in the structure, compared to the snapshots of the full and explicitly solvated system. Kongsted & Ryde have suggested an approach to solve this problem by introducing a buffer region around the truncated system.[8] This buffer region typically contains all protein residues and water molecules within 4 Å of the truncated protein system. It is kept fixed during the minimisation of the system and it is excluded from the frequency calculation. The buffer ensures that there are no major changes in the structure during the minimisation. Thereby, the questionable use of the distance-dependent dielectric constant can also be avoided. Test calculations showed that this simple modification improved the precision of the entropy term by a factor of 2–4, ensuring that it no longer limits the precision of MM/GBSA.[8]

In this article, we present a detailed analysis of the NMA entropy. We investigate the size-dependence of the truncation and study how well truncated systems reproduce calculations with the full protein. Furthermore, we test whether a distance-dependent dielectric constant affects the results. As test cases, we use eight inhibitors of the blood-clotting factor Xa (fXa), nine inhibitors of ferritin, and two ligands of wild-type and a mutant HIV-1 protease. These systems have been used in previous studies of free energy methods.[15,16,17]

Methods

Preparation of proteins and ligands. In this paper, we study eight 3-amidinobenzyl-1*H*-indole-2-carboxamide inhibitors of fXa,[18] nine inhibitors to ferritin, as well as the two inhibitors amprenavir (APV) and darunavir (DRV) binding to either wild-type (WT) or a double mutant (V82T/I84V, MT) HIV-1 protease (HIV-PR). All ligands are shown in Figure 1.

The preparation of the fXa and ferritin ligands has been previously described.[15,16] The fXa simulations were based on the crystal structure of fXa in complex with ligand C125 (PDB id. 1LPK)[18] and the preparation of the protein has been described previously.[8] The ferritin simulations were based on the crystal structure of ferritin in complex with phenol (L09; PDB id. 3f39)[19] and the preparation of the protein has been described previously.[20] The ligands were described by the general Amber force field[21] (GAFF) with charges from restrained electrostatic potentials (RESP)[22] calculations at the HF/6-31G* level and the proteins were described with the Amber99[23] (fXa) or Amber99SB[24] (ferritin) force field. All complexes were solvated in truncated octahedral boxes of TIP3P water molecules, extending at least 10 Å from the protein.[25]

The HIV-PR simulations of WT-APV, MT-APV, WT-DRV, and MT-DRV were based on the corresponding crystal structures with PDB ids. 1HPV, 1T7J, 1T3R, and 1T7I,[26,27] respectively. Crystal water molecules were removed, except one structurally important water molecule between the protease flaps and inhibitor. One of the catalytic Asp residues was protonated (Asp 25), whereas all the other Asp and Glu residues were negatively charged and all Arg and Lys residues were positively charged. The two His residues were protonated on the NE2 atom. The protein was described with the Amber03 force field[28] and the ligands with GAFF. The ligands were optimized at the HF/6-31G** level using Gaussian09.[29] The electrostatic potential (ESP) was then calculated at the B3LYP/cc-pVTZ level with the polarisable continuum model[30] and a dielectric constant of $\varepsilon = 4$ at points sampled with the Merz–Kollman scheme.[31] Point charges were fitted to the ESP using the RESP[22] procedure with the antechamber program. All complexes were immersed in a truncated octahedral box of TIP3P water molecules[25] extending 10 Å from the protein. Finally, between five and seven chlorine ions were added to neutralise the system.

MD simulations. The MD simulations of the fXa and ferritin complexes were conducted by the Amber10 software[32] and they were taken from previous studies.[15,16] The simulations of the HIV-PR complexes were conducted either with the sander module of Amber 10 (equilibration) or the GPU version of the Amber 11 pmemd module (production). The temperature was kept constant at 300 K using a Langevin thermostat[33] and the pressure was kept constant at 1 atm using a weak-coupling algorithm.[34] The non-bonded pair list was updated every 25th step and the non-bonded cut-off was 8 Å (fXa or ferritin) or 9 Å (HIV-PR). Long-range electrostatics and van der Waals interactions were treated using particle-mesh Ewald summation[35] and a continuum approach, respectively. SHAKE constraints[36] were applied to all bonds involving hydrogen atoms and the time step was 2 fs.

The fXa and ferritin complexes were simulated as follows: First, the system was optimised by 500 steps of steepest descent minimisation, keeping all atoms, except water molecules and hydrogen atoms, restrained to their starting positions with a force constant of 418 kJ/mol/Å^2. The minimisation was followed by 20 ps MD equilibration with a constant pressure and the restraining force reduced to 214 kJ/mol/Å^2. Finally, a 1200 ps simulation at a constant pressure, but without any restraints, was performed and the final 200 ps were used for energy calculations. For each complex, we ran 40 independent simulations by assigning different starting velocities to atoms and extracted snapshots every 5 ps so that in total we evaluated MM/GSBA energies on 40 x 40 = 1600 snapshots. This has been shown to be a good protocol on several test cases.[37]

The HIV-PR complexes were simulated as follows: Water molecules and hydrogen atoms

were relaxed first using 100 cycles steepest descent followed by 100 cycles conjugate gradient with all protein and ligand heavy atoms restrained with a force constant of 17 kJ/mol/Å^2. The solute was relaxed by 1000 cycles steepest descent followed by 1000 cycles conjugate gradient without restraints. Initial velocities were randomly assigned from a Boltzmann distribution. The system was then gradually heated from 0 to 300 K over 50 ps and equilibrated for another 100 ps in the NVT ensemble. Finally, 1 ns production simulations were carried out in the NPT ensemble. For each system, equilibration and production steps were repeated 20 times with both different random seeds and different initial water boxes.[38] Snapshots were extracted every fifth picosecond, so that in total we evaluated the MM/GBSA energies on 200 x 20 = 4000 snapshots.

MM/GBSA Calculations. ΔG_{bind} was calculated according to Eqn 1. The terms were calculated with Amber 10 with all water molecules stripped off and without any periodic boundary conditions, but with an infinite cutoff. The MM energies were estimated using the same force field as in the simulations. The polar solvation energy was calculated by the generalised Born (GB) model of Onufriev et al.,[39] model I ($\alpha = 0.8$, $\beta = 0$, and $\gamma = 2.91$). The non-polar solvation energy was estimated from the solvent-accessible surface area (SASA) according to $G_{np} = 0.0227$ SASA (Å^2) + 3.85 kJ/mol.[40]

The translational and rotational entropy was calculated with standard statistical mechanical formulae.[2,7] The vibrational entropies were estimated from the vibrational frequencies using the ideal-gas rigid-rotor harmonic-oscillator approximation.[7] As mentioned in the introduction, the systems were minimised before the frequencies were calculated. Both the minimisation and the frequency calculations were performed in a vacuum.

Several approaches were employed for the entropy calculations, differing in the considered system and the dielectric constant. In one set of calculations, called S_F (full) in the following, the entire complex was considered, but no water molecules. The second set of calculations included the full complex, but also a 4 Å shell of explicit water molecules, surrounding the complex (Figure 2). This buffer was kept fixed during the minimisation and the frequency calculations, and no frequencies and entropies were calculated for this buffer.[8] This approach will be denoted S_R (reference). In the third approach, all protein residues (but no water molecules) within a radius R of the ligand (including the ligand itself) were included in the calculations, together with a buffer that consisted of all residues within 4 Å of the former residues, as well as all water molecules within R + 4 Å of the ligand (Figure 2). Again, the buffer region was kept fixed during the minimisation and the frequency calculation. Radii of $R = 8, 12, 14$, and 16 Å were tested (the radii of the protein–ligand complexes are 35, 40, and 25 Å for fXa, ferritin, and HIV-PR, respectively). These calculations will be denoted by S_{TR}, e.g., S_{T8}. All three types of calculations were performed in two variants viz., with a unit dielectric constant ($\varepsilon = 1$) and with a distance-dependent dielectric constant, $\varepsilon = 4r$, where r is the distance between the involved atoms. The latter calculations will be suffixed by r, e.g., S_{Fr}, S_{Rr}, and S_{T8r}. The six approaches are summarised in Table 1. The protein entropy was calculated with the same methods as for the complex, but with the ligand omitted. The ligand entropy was always calculated for the isolated ligand, but with either $\varepsilon = 1$ or $\varepsilon = 4r$. Throughout this article, all entropies are reported in energy units, i.e. as $-T\Delta S_{bind}$ at 300 K, where $\Delta S_{bind} = S(PL) - S(P) - S(L)$.

Uncertainties and quality metrics. The reported uncertainties for the entropy and free-energy estimates are standard deviation of the averages over the 20 or 40 independent simulations, ignoring the variance within each simulation (i.e. the standard deviation over the independent simulations divided by $\sqrt{20}$ or $\sqrt{40}$).

The quality of the calculated binding affinities compared to experiments is described using the mean unsigned error (MUEtr) after removal of the systematic error (i.e. the mean signed deviation), the correlation coefficient (r^2), and Kendall's rank correlation coefficient. The latter was calculated

after removing pairs of ligands that have either an estimated or experimental difference that is not statistical significant at the 90% confidence level (τ_{90}), because we cannot require that the calculations should reproduce random variations correctly.[41] The number of pairs included in τ_{90} will be denoted by np_{90}. The uncertainties of these metrics were obtained by a simple parametric bootstrap, employing 1000 random samples and assuming normal distribution of the calculated affinities.[37] Uncertainties in both the calculated and experimental values were included in the calculation. As no uncertainty of the experimental binding affinities has been reported, we assumed a typical uncertainty of 1.7 kJ/mol for all three test cases.[42]

Results

Entropies of the fXa complexes. We have computed the MM/GBSA NMA entropy in several different ways (as described in Table 1) for the eight inhibitors of fXa shown in Figure 1. The calculated entropies for this test system are given in Table 2. As a reference, we choose the calculation of the full protein–ligand complex, including a 4 Å shell of explicit water molecules that were kept fixed during the minimisation and the frequency calculations (S_R). These are time-consuming calculations, taking ~2 h per snapshot (i.e., calculations of the complex, protein, and ligand entropies) on a single 3.0 GHz Intel Xenon processor (i.e. 133 CPU days for the results of each ligand in Table 2). Moreover, for larger proteins, the memory requirements may become prohibitively large for typical batch computers (>4 GB). Therefore, the complex is usually truncated. For instance, the S_{T8} calculation takes only ~3 minutes per snapshot.

The results in Table 2 show that S_{T8} deviates on average by 5 kJ/mol from S_R. Ligand C49 shows the largest deviation of 8±2 kJ/mol and most ligands have deviations of ~4 kJ/mol. Interestingly, S_{T8} is always more positive than S_R, showing that the complex and the free protein are affected by the truncation somewhat differently (the ligand entropy is not affected by the truncation). This systematic deviation means that relative entropies (i.e. the difference in entropies among the various ligands) are more accurate than the absolute entropies. In fact, the error in the relative entropies is only 1.3 kJ/mol on average (–1 to 3 kJ/mol). The correlation coefficient (r^2) between S_{T8} and S_R is 0.7 (note that the range of the entropies is only 6–9 kJ/mol).

For all ligands except C50, the deviation decreases going from S_{T8} to S_{T10} (and the difference for C50 is not statistically significant). The mean absolute deviation (MAD) between S_R and S_{T10} is only 3 kJ/mol. With even larger systems, the deviation decreases further, so that the MAD for S_{T12}, S_{T14}, and S_{T16}, the MAD decreases from 2 to 1 kJ/mol, indicating that the calculations converge towards S_R. The maximum deviation decreases somewhat less and is 3 kJ/mol for the three largest truncations. On the other hand, the relative error towards S_R is the same for all truncated systems, 1.3–1.6 kJ/mol.

It is also of interest to compare the reference calculations with the calculations of the full complex without the water buffer (S_F). From Table 2, it can be seen that the MAD is 4 kJ/mol with a maximum deviation of 8 kJ/mol for C125. This indicates that the water buffer stabilises the energies quite strongly: If the buffer is omitted, the entropies change almost as much as for the most truncated system. However, due to the large uncertainties, the difference between S_F and S_R is only significant for three of the ligands at 95% confidence. The standard error is always larger for S_F than for S_R, viz. 2.3 compared to 1.7 kJ/mol on average, and the standard errors for the truncated calculations are even lower, only 0.6 kJ/mol on average for all the truncated calculations, irrespective of the radius. This shows that the truncations stabilise the calculations (decrease the fluctuations among the snapshots), probably because the minimised system is smaller, so there is a smaller number of possible conformations.

Next, we repeated all the calculations using $\varepsilon = 4r$, rather than $\varepsilon = 1$. This can be seen as a primitive but cheap way to add solvation effects to the system. The S_{Rr} entropy does not differ much

from the S_R entropy, 1 kJ/mol on average and with a maximum deviation of 2 kJ/mol ($r^2 = 0.9$), showing that for a full protein with an explicit water buffer, $\varepsilon = 4r$ is a redundant but harmless approximation. The entropies obtained by truncating the protein are more erratic with $\varepsilon = 4r$ than with $\varepsilon = 1$. For three of the ligands, it seems that the entropies converge to within 2 kJ/mol of the S_{Rr} entropy already with S_{T8r}, but for other ligands the deviations compared to S_{Rr} are 4 to 8 kJ/mol and for some ligands, it increases with increasing truncation radius. Consequently, the MAD is almost constant, 3–4 kJ/mol for all truncation radii.

Interestingly, S_{Fr} gives a MAD from S_{Rr} of 12 kJ/mol, which shows that without the buffer, the minimised system can change extensively. Likewise, the MAD between S_F and S_{Fr} is 9 kJ/mol, indicating that using $\varepsilon = 4r$ for a calculation of the full protein changes the structures and hence the entropies significantly. This shows that the commonly used S_{Fr} approach is quite suspicious.

The standard errors using $\varepsilon = 4r$ are slightly larger than with $\varepsilon = 1$, 0.7–0.8 kJ/mol on average for the truncated systems and slightly increasing with the truncation radius. However, for the full systems, the standard errors are similar or slightly smaller with $\varepsilon = 4r$ than with $\varepsilon = 1$, 1.5 kJ/mol for S_{Rr} and 2.2 kJ/mol for S_{Fr} on average.

Free energies of the fXa complexes. Next, we combined the entropies with estimates of interaction energies and solvation free energies to obtain binding free energies according to Eqn 1. These binding free energies are shown in Table 3. The trends due to entropy estimates have already been discussed in the previous section, so here the interesting thing is to compare the calculated binding affinities with the experimental data (shown in Figure 1).[18] It can be seen that the reference entropy estimates (S_R) give reasonably good results: The MUEtr is 3±1 kJ/mol, which is slightly better than the null hypothesis that all estimates are equal, with MUEtr = 5 kJ/mol (reflecting that all the experimental binding affinities are quite similar; in fact six of them are within 5 kJ/mol). The correlation coefficient is rather good, $r^2 = 0.7\pm0.1$ and Kendall's rank correlation coefficient, based on the pairs that have statistically significant differences in both the experiments and calculations (12 out of 28 possible pairs) is perfect, $\tau_{90} = 1.0\pm0.1$.

From Table 3, it can be seen that all the other entropy methods give similar results: MUEtr = 2–4 kJ/mol, with differences that are not statistically significant at the 95% level. However, all calculations with $\varepsilon = 4r$ give a higher MUEtr than the corresponding calculations with $\varepsilon = 1$. Likewise, differences in r^2 (0.4–0.8) and τ_{90} (0.8–1.0) are not statistically significant. The latter two measures show no clear difference between the two dielectric constants. The reason why we do not see any clear differences between the binding affinities obtained with the various entropies is that the relative entropies vary by only ~1.5 kJ/mol for S_{TR} and by 2–3 kJ/mol for S_{TRr}. This variation is within the uncertainty of the method (note we use quality measures that are sensitive only to the relative affinities, not to the absolute ones, because the absolute MM/GBSA affinities depend strongly on the continuum-solvation model and rarely give accurate absolute affinities;[43,44] moreover absolute affinities are seldom of interest in drug development). However, it is notable that S_F gives the poorest correlation coefficient ($r^2 = 0.4$) and S_{Fr} gives the poorest $\tau_{90} = 0.8$ among all tested entropy methods – these two approaches also gave entropies that differed most from S_R.

Moreover, it is clear that entropies calculated with the full protein had a higher uncertainty than those obtained for the truncated systems and that this uncertainty dominates the total binding affinities: The standard error of ΔG_{bind} obtained with S_R is 2 kJ/mol, whereas it is only 1 kJ/mol when obtained with S_{TR} for any radius. This is an important advantage of the entropies obtained with truncated systems.

Owing to the high cost of the entropy calculations in MM/GBSA, the entropy term has in several cases been completely ignored.[45,46,47,48] We have also tested such an approach for fXa and present the corresponding affinity estimates in Table 4. It can be seen that the omission of the entropy term makes all binding affinities much too negative (−165 to −182 kJ/mol, compared to the

experimental estimates of –27 to –47 kJ/mol). However, it can be seen that the MUEtr (3 kJ/mol), r^2 (0.7), and τ_{90} (1.0) are similar to what is obtained for MM/GBSA with the various entropy estimates. Hence, for this test case, the entropy estimates do not significantly change the relative affinities.

Ferritin and HIV-1 protease. From the calculations on the fXa complexes, it seems sufficient to calculate the entropies with S_{T8} if we are only interested in relative energies or entropies. To test this hypothesis we performed S_R and S_{T8} calculations also for the binding of nine inhibitors to ferritin and for the binding of two ligands to two different forms of HIV-PR. For ferritin, it can be seen in Table 5 that S_{T8} is an excellent approximation to S_R with deviations of 0.1 to 1.7 kJ/mol, giving a MAD of only 0.9 kJ/mol for absolute entropies and 0.7 kJ/mol for relative entropies. Consequently, there are no significant differences in the binding free energies calculated with the two entropy approaches. They both give a MUEtr of 9 kJ/mol compared to the experimental estimates.[19] This is appreciably worse than the null hypothesis that all ligands have the same affinity (giving a MAD of 3 kJ/mol), which reflects that the MM/GBSA method overestimates the difference in affinity between the various ligands. On the other hand, MM/GBSA gives excellent r^2 = 0.9 and τ_{90} = 1.0 (based on 20 out of the 36 possible pairs). Omitting the entropy estimate completely gives a significantly worse MUEtr (14 kJ/mol), but similar r^2 and τ_{90}. The precision of the two entropy estimates and the binding free energies is improved for the truncated systems, e.g. from 1.3 to 0.6 kJ/mol on average for the entropy. It should be mentioned that for this test case, the molecular weights correlates well with the experimental affinities (r^2 = 0.86). However, this is due to the large range of weights, from 206.3 g/mol for L02 to 94.1 g/mol for L09 (phenol) and that the interaction between drug and receptor is mainly dictated by van der Waals forces. The large range of weights leads to a MADtr of 33.3 kJ/mol, i.e., much worse than any of the MM/GBSA estimates.

For the HIV-PR complexes, the difference between S_{T8} and S_R ranges from 3 to 5 kJ/mol, as can be seen in Table 6, giving a MAD of 4 kJ/mol. However, the difference is systematic so the difference in relative entropies is only 0.8 kJ/mol, which shows that relative entropies can accurately be calculated with S_{T8} also for this test case. Consequently, the two entropy approaches give similar binding affinities. The MUEtr of ΔG_{bind} compared to experimental data[49] is 4–5 kJ/mol, which is similar to the null hypothesis (5 kJ/mol). The r^2 is rather low (0.45–0.55) but τ_{90} = 1.0. The entropy difference between WT and MT for APV, 1–3 kJ/mol reproduces the isothermal calorimetry value of 2.5 kJ/mol.[49] The corresponding difference for DRV, 0 to –2 kJ/mol, is also reasonably close to the experimental value of –2.5 kJ/mol. Omitting, the entropy, gives a somewhat worse MUEtr, the same τ_{90}, and a slightly better r^2, but none of the differences is statistically significant at the 90% level. Interestingly, the precision of the entropy is not so much improved for the truncated protein as for the other test cases.

Conclusions
In this article, we have investigated the effect of truncation and the dielectric constant on NMA entropy estimates, employed in the MM/GBSA method. As the reference, we use the full protein surrounded with a shell of fixed water molecules (S_R). As test cases, we study the binding of eight inhibitors to fXa, nine ligands to ferritin, and two ligands to wild-type and mutant HIV-PR. Several interesting results are obtained.

- Truncation of the protein is not innocent: Truncating the protein outside a radius of 8–16 Å from the ligand gives rise to average deviations of 1–5 kJ/mol, and the deviations increase when the radius is decreased. Fortunately, the deviations are systematic, so for relative entropies, the average deviation is only 0.7–1.6 kJ/mol for all radii. Consequently, we do not see any statistically significant difference for the total ligand-binding affinities with any of our quality measures (which compare only relative affinities) when the systems are truncated. On the other hand, the precision is nearly always improved for the truncated

systems.

- Calculations with a distance-dependent dielectric constant ($\varepsilon = 4r$, a primitive model to account for solvent effects when water molecules are excluded) give a rather constant MAD from S_{Rr} of 3–4 kJ/mol, independent from the truncation radius. This is similar to the results with $\varepsilon = 1$ for the smallest radii. However, the deviations are less systematic, so for relative entropies, the average deviations are larger than with $\varepsilon = 1$, 2–3 kJ/mol. Entropies calculated with $\varepsilon = 4r$ differ by 1–2 kJ/mol on average from those calculated with $\varepsilon = 1$.

- A shell of fixed water molecules and protein residues for the truncated systems is important to stabilise the energies. Without this shell, the entropies change by 4 kJ/mol on average with $\varepsilon = 1$ and by 12 kJ/mol with $\varepsilon = 4r$. Moreover, the standard error increases by 0.6 kJ/mol on average.

- We have tested to ignore the entropy term completely in MM/GBSA. Such an approach has been shown to improve the results in several studies,[45,46,47,48] although the results become worse in other cases.[16,46] For the fXa and HIV–PR test cases, it does not lead to any statistically significant changes in our quality measures, but for ferritin, the results become slightly worse. Thus, omission of the entropy term is not always favourable or innocent.[46]

Of course, our results depend on the selection on S_R as the reference entropy. Other references are conceivable, but not necessarily better. The most natural would be the full complex with all water molecules, treated the same way as in the MD simulations, i.e. with periodic boundary conditions and Ewald summation. Unfortunately, this leads to too large systems (the memory requirements become prohibitive) to treat with today's computers. Moreover, the water molecules need to be excluded from the frequency calculations (the water entropy is included in the two solvation terms) and a minimisation of unconstrained systems may lead to extensive changes in the structure. For the latter reason, we believe that a fixed shell of water molecules and protein residues is the best solution to keep the structure close to the MD snapshot. It also leads to an improved precision. Once, such a buffer is used, the need of solvation effects becomes minor, as the similarity of the results with $\varepsilon = 1$ and $\varepsilon = 4r$ show. Therefore we do not expect large improvement by using more sophisticated continuum-solvation methods as e.g. generalised Born, as has recently been suggested.[12,50] We favour the approach proposed here as it retains a quite consistent energy model between the underlying MD simulations and the subsequent estimation of free energy changes, and is computationally more efficient due to the smaller size of the truncated system.

Thus, we can conclude that entropies calculated with a protein truncated beyond a distance of 8 Å from the ligand and with a 4 Å buffer of fixed protein residues and water molecules, together with $\varepsilon = 1$[8] provide a proper compromise between accuracy and computational efficiency, albeit only for relative binding affinities. However, it has been repeatedly pointed out that MM/GBSA cannot be expected to give accurate absolute affinities, only relative affinities.[43,44,51] Finally, it should also be remembered that NMA is a crude approximation because it assumes that phase space is Gaussian-shaped with a single well, although it is clear that many degrees of freedom, in particular many dihedral angles, have multiple minima. Therefore, other methods to calculate the entropy may give significantly different results.[9,11]

References

1. Srinivasan, J.; Cheatham III, T. E.: Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices *J. Am. Chem. Soc.* **1998,** *120,* 9401-9409

2. Kollman, P.; A.; 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, D. A.; Cheatham, T. E., III. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models *Acc. Chem. Res.* **2000,** *33,* 889– 897

3. Foloppe N.; Hubbard, R. Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* **2006** *13,* 3583-3608

4. Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy *Biophys. J.* **2004**, *86,* 67– 74

5. Pearlman, D. A. Evaluating the Molecular Mechanics Poisson−Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to p38 MAP Kinase *J. Med. Chem.* **2005**, *48,* 7796– 7807

6. Genheden, S.; Ryde, U. Comparison of end-point continuum-solvation methods for the calculation of protein–ligand binding free energies. *Proteins*, **2012** *80,* 1326–1342

7. Jensen F. *Introduction to computational chemistry.* **1999** Wiley, Chichester

8. Kongsted, J.; Ryde, U. An Improved Method to Predict the Entropy Term with the MM/PBSA Approach *J. Comput-Aided Mol. Design.* **2009**, *23,* 63– 71

9. Diehl, C.; Genheden, S.; Modig, K.; Ryde, U.; Akke, M. Conformational entropy changes upon lactose binding to the carbohydrate recognition domain of galectin-3 J. Biomol. NMR., **2009**, *45,* 157-169

10. Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein−Protein Complex Ras−Raf *J. Comput. Chem.* **2003**, *25,* 238– 250

11. Genheden, S., Ryde, U. Will molecular dynamics simulations of proteins ever reach equilibrium? *Phys. Chem. Chem. Phys.*, **2012**, *14,* 8662-8677.

12. Kopitz, H.; Cashman, D. A.; Pfeiffer-Marek, S.; Gohlke, H. Influence of the Solvent Representation on Vibrational Entropy Calculations: Generalized Born Versus Distance-Dependent Dielectric Model. *J. Comput. Chem.* **2012**, *33,* 1004-1013

13. Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. Ligand Affinities Predicted with the MM/PBSA Method: Dependence on the Simulation Method and the Force Field *J. Med. Chem.* **2006**, *49,* 6596– 6606

14. Page, C.S.; Bates, P. A. Can MM-PBSA calculations predict the specificities of protein kinase inhibitors? *J. Comput. Chem.* **2006**, *27,* 1990-2007

15. Genheden, S.; Nilsson, I.; Ryde, U. Binding Affinities of Factor Xa Inhibitors Estimated by Thermodynamic Integration and MM/GBSA. *J. Chem. Inf. Model.* **2011**, *51,* 947–958.

16. Mikulskis, P.; Genheden, S.; Wichmann, K.; Ryde, U. A semiempirical approach to ligand-binding affinities: Dependence on the Hamiltonian and corrections. *J. Comput. Chem.* **2012**, *33,* 1179-1189

17. Sadiq, S. K.; Wright, D. W.; Kenway, O. A.; Coveney, P. V. Accurate Ensemble Molecular Dynamics Binding Free Energy Ranking of Multidrug-Resistant HIV-1 Proteases *J. Chem. Inf. Model.* **2010**, *50,* 890– 905

18. Matter, H.; Defossa, E.; Heinelt, U.; Blohm, P.-M.; Schneider, D.; Müller, A.; Herok, S. I.; Schreuder, H.; Liesum, A.; Brachvogel, V.; Lönze, P.; Walser, A.; Al-Obeidi, F.; Wildgoose, P. Design and Quantitative Structure-Activity Relationship of 3-Amidinobenzyl-1H-indole-2-carboxamides as Potent, Nonchiral, and Selective Inhibitors of Blood Coagulation Factor Xa *J. Med. Chem.* **2002**, *45,* 2749– 2769

19. Vedula, L. S.; Brannigan, G.; Economou, N. J.; Xi, J.; Hall, M. A.; Liu, R.; Rossi, M. J.; Dailey, W. P.; Grasty, K. C.; Klein, M. L; Eckenhoff, R. G.; Loll, P. J. Protein Structure and Folding. *J. Bio. Chem.* **2009**, *284,* 24176-24184.

20. Genheden, S.; Mikulskis, P.; Hu, L.; Kongsted, J.; Söderhjelm, P.; Ryde, U. Accurate estimation of non-polar solvation free energies requires explicit consideration of binding site hydration *J. Am. Chem. Soc.*, **2011**, *133,* 13081-1309

21. Wang, J. M.; Wolf, R. M.; Caldwell, K. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field *J. Comput. Chem.* **2004**, *25,* 1157– 1174

22. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model *J. Phys. Chem.* **1993**, *97,* 10269– 10280

23. Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21,* 1049-1074

24. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct. Bioinform.* **2006**, *65,* 712

25. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impley, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water *J. Chem. Phys.* **1983**, *79,* 926– 935

26. Kim, E. E.; Bake, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.* **1995**, *117,* 1181-1182

27. Surleraux, D. L. N. G.; Tahri, A.; Verschueren, W. G.; Pille, G. M. E.; de Kock, H. A.; Jonckers, T. H. M.; Peeters,

A.; De Meyer, S.; Azijn, H.; Pauwels, R.; de Bethune, M-P.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B. T. P. Discovery and Selection of TMC114, a Next Generation HIV-1 Protease Inhibitor. *J. Med. Chem.*, **2005**, *48*, 1813-1822

28 Duan, Y.; Wu, C.; Shibasish, C.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.: Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999-2012

29 Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision **A.1**, Gaussian, Inc., Wallingford CT, **2009**.

30 Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B*, **2009**, *113*, 6378-96.

31 Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods *J. Comput. Chem.* **1990**, *11*, 431– 439

32 Case, D. A.; Darden, T. A.; Cheatham, T. E., III; 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**.

33 Wu, X.; Brooks, B. R. Self-Guided Langevin Dynamics Simulation Method *Chem. Phys. Lett.* **2003**, *381*, 512– 518

34 Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath *J. Chem. Phys.* **1984**, *81*, 3684– 3690

35 Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems *J. Chem. Phys.* **1993**, *98*, 10089– 10092

36 Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes *J. Comput. Phys.* **1977**, *23*, 327– 341

37 Genheden, S.; Ryde, U. How To Obtain Statistically Converged MM/GBSA Results *J. Comput. Chem.* **2010**, *31*, 837– 846

38 Genheden, S.; Ryde, U. A Comparison of Different Initialisation Protocols to Obtain Statistically Independent Molecular Dynamics Simulations *J. Comput. Chem.* **2011**, *32*, 187– 195

39 Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model *Proteins* **2004**, *55*, 383– 394

40 Kuhn, B.; Kollman, P. A. Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* **2000**, *43*, 3786-3791

41 Mikulskis, P.; Genheden, S.; Rydberg, P.; Sandberg, L.; Olsen, L.; Ryde, U. Binding affinities in the SAMPL3 trypsin and host–guest blind tests estimated with the MM/PBSA and LIE methods *J. Comp.-Aided Mol. Des.*, **2012**, in press; 10.1007/s10822-011-9524-z.

42 Brown, S. P.; Muchmore, S. W.; Hajduk, P. J. Healthy skepticism: assessing realistic model performance. *Drug Discov. Today* **2009,** *14*, 420-427

43 Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. An MM/3D-RISM Approach for Ligand-Binding Affinities *J. Phys. Chem. B.* **2010**, *114*, 8505– 8516

44 Singh, N.; Warshel, A. Absolute binding free energy calculations: on the accuracy of computational scoring of protein-ligand interactions. *Proteins* **2010**, *78*, 1705-1723

45 Hayes, J. M.; Skamnaki, V. T.; Archontis, G.; Lamprakis, C.; Sarrou, J.; Bischler, N.; Skaltsounis, A-L.; Zographos, S. E.; Oikonomakos, N. H. Kinetics, in silico docking, molecular dynamics, and MM-GBSA binding studies on prototype indirubins, KT5720, and staurosporine as phosphorylase kinase ATP-binding site inhibitors: the role of water molecules examined. *Proteins* **2011**, *79*, 703-719

46 Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69-82

47 Genheden, S. MM/GBSA and LIE estimates of host-guest affinities: dependence on charges and solvation model. *J.*

Comput-Aided Mol. Des. **2011**, *25*, 1085–1093

48 Yang, T.; Wu, J. C.; Yan, C.; Wang, Y.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Virtual screening using molecular simulations. *Proteins* **2011**, *79*, 1940-1951.

49 King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Wigerinck, P.; Bethune, M.P.; Schiffer, C. A. Structural and Thermodynamic Basis for the Binding of TMC114, a Next-Generation Human Immunodeficiency Virus Type 1 Protease Inhibitor. *J. Virol.* **2004,** *78,* 12012-12021

50 Brown, R A.; Case, D. A Second derivatives in generalized Born theory *J. Comput. Chem.* **2006**, *27*, 1662-1675

51 Kongsted, J.; Söderhjelm, P.; Ryde, U. How accurate are continuum solvation models for drug-like molecules? *J. Comp.-Aided Mol. Des.* **2009**, *23*, 395-409

Table 1. Summary of entropy methods employed in this study.[a]

Abbreviation	Residues	Residue buffer	Water buffer	Dielectric constant
S_F	All	No	No	1
S_R	All	No	Yes	1
S_{TR}	All within R Å	All between R and $R + 4$	Yes	1
S_{Fr}	All	No	No	$4r$
S_{Rr}	All	No	Yes	$4r$
S_{TRr}	All within R Å	All between R and $R + 4$	Yes	$4r$

[a] $R = 8, 10, 12, 14,$ and 16 Å. The residue and water buffer is kept fixed during the minimisation and the frequency calculation. No frequencies or entropies are calculated for the buffer atoms.

Table 2. Total entropy ($-T\Delta S_{\text{bind}}$) for the binding of the eight inhibitors to fXa (kJ/mol) calculated with 14 different approaches (explained in Table 1).

Method	S_R	S_F	S_{T8}	S_{T10}	S_{T12}	S_{T14}	S_{T16}
C9	113.3 ±1.9	116.1 ±2.1	117.7 ±0.6	113.0 ±0.6	110.9 ±0.6	110.9 ±0.7	110.3 ±0.6
C39	112.5 ±1.4	107.7 ±2.3	116.5 ±0.5	114.1 ±0.7	115.0 ±0.7	114.4 ±0.5	114.5 ±0.6
C47	113.7 ±1.6	116.2 ±2.0	117.5 ±0.7	115.6 ±0.6	113.2 ±0.6	113.2 ±0.5	112.9 ±0.6
C49	112.9 ±1.7	119.7 ±2.4	121.0 ±0.6	117.7 ±0.6	114.1 ±0.6	116.1 ±0.6	115.0 ±0.6
C50	113.1 ±1.5	115.9 ±2.4	117.1 ±0.7	117.8 ±0.6	115.3 ±0.6	114.2 ±0.7	114.2 ±0.7
C53	110.4 ±1.7	116.5 ±2.1	116.4 ±0.5	111.4 ±0.6	111.2 ±0.6	110.3 ±0.6	109.6 ±0.6
C63	107.6 ±1.8	109.2 ±2.2	111.6 ±0.5	110.5 ±0.6	111.1 ±0.6	110.2 ±0.6	108.9 ±0.7
C125	107.4 ±1.6	115.8 ±2.7	113.5 ±0.6	111.6 ±0.6	109.2 ±0.7	109.0 ±0.6	108.2 ±0.6
MAD[a]		4.5	5.0	2.7	1.9	1.7	1.5

Method	S_{Rr}	S_{Fr}	S_{T8r}	S_{T10r}	S_{T12r}	S_{T14r}	S_{T16r}
C9	111.6 ±1.4	124.5 ±2.3	113.5 ±0.7	110.0 ±0.8	110.0 ±0.8	110.9 ±0.8	111.7 ±0.8
C39	110.9 ±1.6	126.4 ±2.3	117.7 ±0.7	117.6 ±0.8	118.7 ±0.7	118.9 ±0.7	119.1 ±0.8
C47	112.1 ±1.7	128.3 ±2.7	115.4 ±0.8	114.1 ±0.8	113.0 ±0.7	113.3 ±0.8	114.2 ±0.8
C49	112.2 ±1.6	122.0 ±2.3	120.1 ±0.8	117.4 ±0.8	110.5 ±0.7	118.0 ±0.9	118.3 ±0.9
C50	113.8 ±1.4	125.4 ±2.3	115.3 ±0.6	116.8 ±0.9	116.4 ±0.8	116.4 ±0.7	117.0 ±0.7
C53	109.2 ±1.6	119.9 ±2.3	111.4 ±0.7	109.7 ±0.7	110.4 ±0.7	110.0 ±0.7	110.5 ±0.8
C63	106.8 ±1.3	121.5 ±2.2	111.0 ±0.6	111.7 ±0.7	112.8 ±0.7	114.3 ±0.7	113.1 ±0.8
C125	107.6 ±1.5	106.8 ±1.3	109.3 ±0.7	107.7 ±0.7	106.3 ±0.7	108.0 ±0.8	108.1 ±0.8
MAD[b]		11.5	3.7	3.0	2.9	3.4	3.5

[a] Mean absolute deviation compared to S_R
[b] Mean absolute deviation compared to S_{Rr}

Table 3. MM/GBSA free energies for the binding of the eight inhibitors to fXa (ΔG_{bind} in kJ/mol) calculated with the various entropy estimates. The results are compared to experimental data[18] using three quality estimates, MUEtr, r^2, and τ_{90}.

Entropy	S_R	S_F	S_{T8}	S_{T10}	S_{T12}	S_{T14}	S_{T16}	Exp[18]
C9	-68.6 ±2.1	-65.8 ±2.2	-64.2 ±1.0	-68.9 ±1.0	-71.0 ±0.9	-71.0 ±1.0	-71.6 ±0.9	-46.2
C39	-53.3 ±1.7	-58.1 ±2.0	-49.3 ±1.0	-51.7 ±1.0	-50.8 ±1.0	-51.5 ±1.0	-51.3 ±1.1	-27.3
C47	-63.0 ±2.0	-60.5 ±2.3	-59.2 ±1.1	-61.1 ±1.0	-63.6 ±1.0	-63.5 ±0.9	-63.8 ±1.0	-46.8
C49	-65.5 ±2.0	-58.7 ±2.5	-57.4 ±1.1	-60.7 ±1.0	-64.3 ±1.1	-62.2 ±1.1	-63.3 ±1.0	-41.9
C50	-63.7 ±1.5	-60.9 ±2.5	-59.7 ±0.9	-58.9 ±0.8	-61.5 ±0.9	-62.6 ±0.9	-62.6 ±0.9	-46.2
C53	-67.3 ±1.9	-61.1 ±2.0	-61.2 ±1.0	-66.3 ±1.1	-66.4 ±1.1	-67.4 ±1.0	-68.0 ±1.1	-44.3
C63	-57.7 ±2.2	-56.2 ±2.5	-53.7 ±1.6	-54.8 ±1.5	-54.3 ±1.5	-55.1 ±1.6	-56.4 ±1.6	-37.4
C125	-69.9 ±1.9	-61.5 ±2.8	-63.8 ±1.2	-65.7 ±1.1	-68.1 ±1.1	-68.3 ±1.0	-69.1 ±1.1	-43.4
MUEtr	2.9 ±0.7	3.3 ±0.8	2.4 ±0.6	3.5 ±0.6	3.4 ±0.7	3.2 ±0.7	3.1 ±0.7	
r^2	0.68 ±0.14	0.39 ±0.18	0.77 ±0.10	0.60 ±0.11	0.70 ±0.19	0.71 ±0.10	0.71 ±0.10	
τ_{90}	1.00 ±0.06	1.00 ±0.16	1.00 ±0.06	0.87 ±0.05	1.00 ±0.05	1.00 ±0.05	1.00 ±0.04	
np_{90}	12	3	15	14	15	14	14	

Entropy	S_{Rr}	S_{Fr}	S_{T8r}	S_{T10r}	S_{T12r}	S_{T14r}	S_{T16r}	Exp[18]
C9	-70.3 ±1.5	-57.4 ±2.3	-68.4 ±0.9	-71.9 ±1.0	-71.9 ±1.0	-71.0 ±1.0	-70.2 ±1.0	-46.2
C39	-54.9 ±1.7	-39.4 ±2.2	-48.2 ±1.0	-48.2 ±1.2	-47.2 ±1.1	-46.9 ±1.1	-46.8 ±1.1	-27.3
C47	-64.6 ±1.7	-48.5 ±2.6	-61.3 ±1.2	-62.6 ±1.1	-63.7 ±1.1	-63.5 ±1.2	-62.5 ±1.2	-46.8
C49	-66.2 ±1.9	-56.4 ±2.4	-58.2 ±1.1	-61.0 ±1.1	-67.9 ±1.1	-60.4 ±1.2	-60.1 ±1.1	-41.9
C50	-63.0 ±1.5	-51.4 ±2.1	-61.5 ±0.9	-60.0 ±1.1	-60.4 ±1.1	-60.4 ±1.0	-59.8 ±1.1	-46.2
C53	-68.4 ±1.7	-57.7 ±2.4	-66.3 ±1.1	-68.0 ±1.1	-67.3 ±1.2	-67.6 ±1.0	-67.1 ±1.1	-44.3
C63	-58.5 ±1.9	-43.8 ±2.2	-54.3 ±1.4	-53.7 ±1.5	-52.5 ±1.5	-51.1 ±1.6	-52.2 ±1.6	-37.4
C125	-69.7 ±1.8	-56.6 ±2.0	-68.0 ±1.0	-69.6 ±1.0	-71.0 ±1.0	-69.3 ±1.1	-69.2 ±1.0	-43.4
MUEtr	3.1 ±0.7	4.0 ±0.9	3.3 ±0.7	3.9 ±0.7	4.5 ±0.7	3.8 ±0.7	3.7 ±0.7	
r^2	0.65 ±0.13	0.58 ±0.14	0.72 ±0.10	0.66 ±0.10	0.65 ±0.10	0.71 ±0.10	0.70 ±0.10	
τ_{90}	1.00 ±0.06	0.83 ±0.05	1.00 ±0.06	1.00 ±0.04	0.75 ±0.04	1.00 ±0.05	1.00 ±0.04	
np_{90}	13	12	16	14	16	15	14	

Table 4. MM/GBSA free energies for the binding of the eight inhibitors to fXa (ΔG_{bind} in kJ/mol) calculated without any entropy estimate.

	ΔG_{bind}	Exp[18]
C9	-181.9 ±0.7	-46.2
C39	-165.8 ±1.0	-27.3
C47	-176.7 ±0.9	-46.8
C49	-178.4 ±0.8	-41.9
C50	-176.8 ±0.9	-46.2
C53	-177.6 ±0.8	-44.3
C63	-165.3 ±1.5	-37.4
C125	-177.3 ±0.9	-43.4
MUEtr	2.9 ±0.6	
r^2	0.71 ±0.11	
τ_{90}	1.00 ±0.04	
np_{90}	13	

Table 5. Entropies and free energies for the binding of nine inhibitors to ferritin calculated with the S_R and S_{T8} approaches, as well as without any entropy estimate (kJ/mol).

	$-T\Delta S_{bind}$		ΔG_{bind}			Exp[19]
	S_R	S_{T8}	S_R	S_{T8}	No S	
L01	67.5 ±1.3	66.7 ±0.6	-37.2 ±1.5	-38.0 ±0.9	-104.7 ±0.6	-30.5
L02	72.4 ±1.4	73.5 ±0.7	-47.5 ±1.7	-46.5 ±1.2	-119.9 ±0.9	-30.1
L03	70.0 ±1.4	69.8 ±0.8	-42.6 ±1.5	-42.7 ±1.0	-112.6 ±0.6	-32.7
L04	68.8 ±1.5	68.4 ±0.6	-37.4 ±1.7	-37.8 ±0.8	-106.2 ±0.6	-30.6
L05	63.1 ±1.2	63.0 ±0.6	-30.4 ±1.4	-30.5 ±0.8	-93.5 ±0.8	-28.2
L06	61.4 ±1.3	60.0 ±0.4	-20.5 ±1.4	-21.9 ±0.6	-81.9 ±0.5	-26.0
L07	60.5 ±1.3	59.5 ±0.5	-20.1 ±1.4	-21.1 ±0.8	-80.6 ±0.6	-27.5
L08	57.4 ±1.2	56.1 ±0.5	-16.0 ±1.2	-17.3 ±0.7	-73.4 ±0.6	-22.8
L09	52.8 ±1.2	51.1 ±0.4	0.9 ±1.4	-0.8 ±1.0	-51.9 ±1.0	-18.7
MUEtr		0.9[a]	9.1 0.5	8.5 0.3	14.2 0.2	
r^2			0.89 0.02	0.90 0.01	0.89 0.01	
τ_{90}			1.00 0.06	1.00 0.05	1.00 0.05	
np_{90}			20	20	20	

[a] Mean absolute difference from the S_R estimates.

Table 6. Entropies and free energies for the binding of two ligands to WT and MT-HIV-PR, calculated with the S_R and S_{T8} approaches, as well as without any entropy estimate (kJ/mol).

	$-T\Delta S_{bind}$		ΔG_{bind}			Exp[49]
	S_R	S_{T8}	S_R	S_{T8}	No S	
MT-APV	101.3 ±0.7	104.8 ±0.6	-151.6 ±2.7	-148.1 ±2.7	-252.9 ±2.6	-49.0
WT-APV	102.3 ±1.1	107.4 ±0.5	-144.1 ±2.0	-139.0 ±1.7	-246.4 ±1.6	-52.8
MT-DRV	113.6 ±1.1	116.7 ±1.0	-158.5 ±2.4	-155.3 ±2.3	-272.0 ±2.1	-57.4
WT-DRV	111.6 ±0.9	116.3 ±0.9	-160.5 ±2.1	-155.8 ±2.1	-272.1 ±1.9	-63.6
MUEtr		4.1[a]	3.9 ±1.0	4.7 ±1.0	6.4 ±1.2	
r^2			0.55 ±0.18	0.45 ±0.16	0.66 ±0.15	
τ_{90}			1.00 ±0.06	1.00 ±0.08	1.00 ±0.09	
np_{90}			4	4	4	

[a] Mean absolute difference from the S_R estimates.

Figure 1. Ligands used in this study. Et, iPr, and sBu indicate ethyl, isopropyl, and secondary butyl groups. The experimental affinities are shown in kJ/mol.[18,19,49]

C9   -46.2

C39   -27.3

C47   -46.8

C50   -41.9

C49   -46.2

C53   -44.3

C63   -37.4

C125   -43.4

L01   -30.5

L02   -30.1

L03   -32.7

L04   -30.6

L05   -28.2

L06   -26.0

L07   -27.5

L08   -22.8

L09   -18.7

Amprenavir   MT: -49.0, WT: -52.8
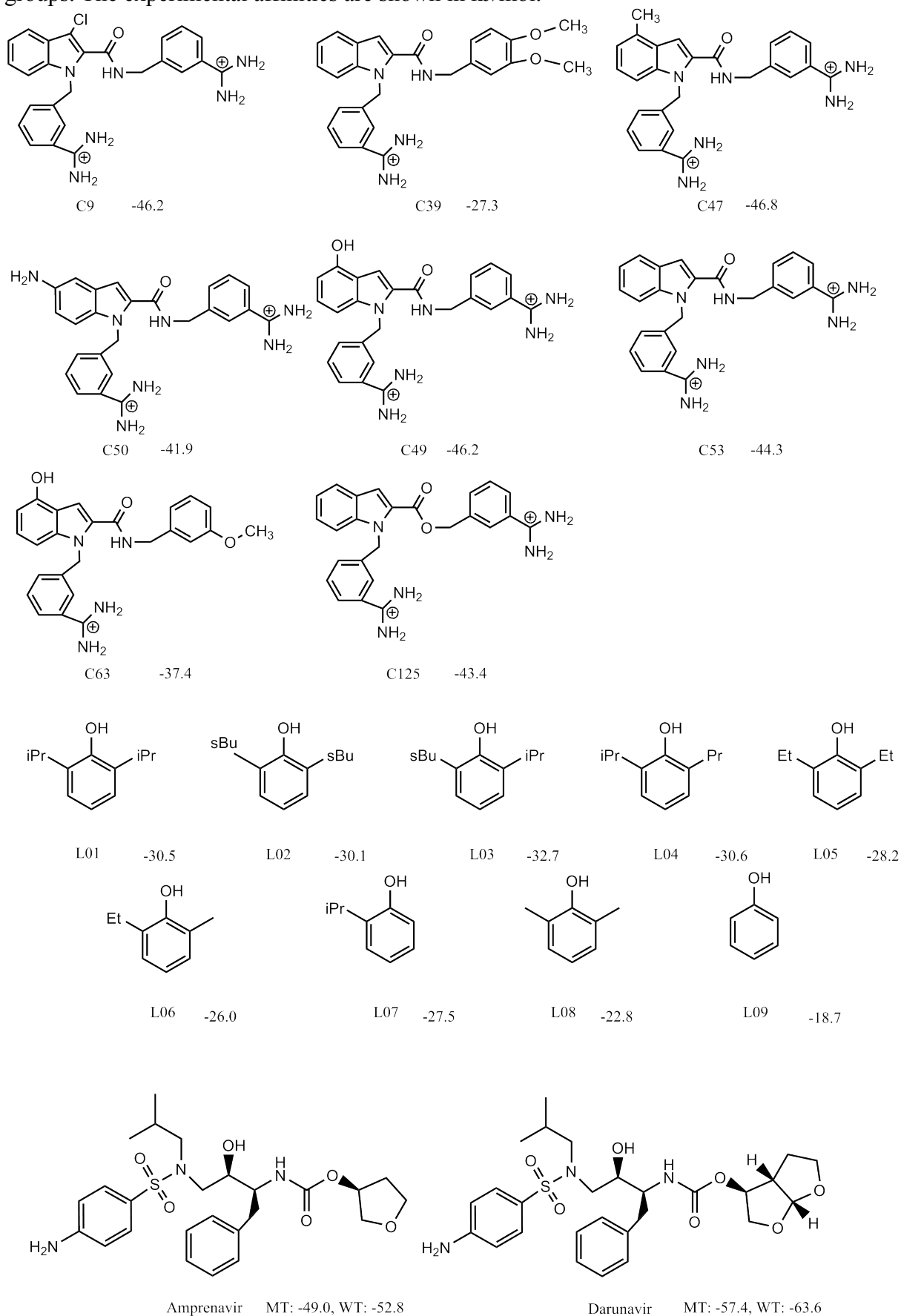
Darunavir   MT: -57.4, WT: -63.6

Figure 2. Illustration of the truncation approaches. Water molecules are shown in blue, protein residues that are included in the frequency calculations in green, and protein residues that are in the buffer region are shown in orange. The ligand is shown in ball-and-stick representation. The protein is fXa with C53.