An improved method to predict the entropy term with the MM/PBSA approach.

Kongsted, Jacob; Ryde, Ulf

Published in:
Journal of Computer-Aided Molecular Design

DOI:
10.1007/s10822-008-9238-z

2009

Document Version:
Peer reviewed version (aka post-print)

An improved method to predict the entropy term with the MM/PBSA approach

Jacob Kongsted & Ulf Ryde *

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

Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se, Tel: +46 – 46 2224502, Fax: +46 – 46 2224543

2017-03-18

Abstract

A method is suggested to calculate improved entropies within the MM/PBSA approach (molecular mechanics combined with Poisson–Boltzmann and surface area calculations) to estimate protein–ligand binding affinities. In the conventional approach, the protein is truncated outside ~8 Å from the ligand. This system is freely minimised using a distance-dependent dielectric constant (to simulate the removed protein and solvent). However, this can lead to extensive changes in the molecular geometry, giving rise to a large standard deviation in this term. In our new approach, we introduce a buffer region ~4 Å outside the truncated protein (including solvent molecules) and keep it fixed during the minimisation. Thereby, we reduce the standard deviation by a factor of 2–4, ensuring that the entropy term no longer limits the precision of the MM/PBSA predictions. The new method is tested for the binding of seven biotin analogues to avidin, eight amidinobenzyl-indole-carboxamide inhibitors to factor Xa, and two substrates to cytochrome P450 3A4 and 2C9. It is shown that it gives more stable results and often improved predictions of the relative binding affinities.

Key Words: entropy, MM/PBSA, avidin, factor Xa, cytochrome P450.

Introduction

One of the major goals of computational chemistry is to develop methods to accurately predict the binding energy of a ligand to a protein. This is of central interest in medicinal chemistry, because the action of most drugs (inhibition, activation, etc.) is caused by the binding of the drug to its target receptor. However, many biochemical problems can be treated in a similar way. For example, the reactivity of an enzyme can be estimated by comparing the free energy of the reactant and transition states of the active site in the protein. Therefore, many methods have been developed with this aim [1,2].

The most accurate and stringent theoretical methods to predict ligand affinities is free energy perturbation (FEP) [3]. In this method, a free energy change is calculated by slowly changing one system into another via a set of unphysical mixed states, using molecular dynamics (MD) or Monte Carlo simulations. Unfortunately, the results converge only for small changes. Therefore, this method has mainly been used to calculate the relative binding affinities of similar drugs to the same protein [Error: Reference source not found,Error: Reference source not found,4]. In addition, more approximate methods to estimate ligand affinities have been developed, e.g. the linear interaction energy (LIE) method [5] and the molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) method [6]. Both methods restrict the molecular simulations to the states before and after the binding process.

The MM/PBSA approach is attractive because it does not contain any parameters that vary for different ligand–receptor systems and it involves a set of physically well-defined terms: The binding affinity is estimated from the free energies of the receptor (R), the ligand (L), and the complex (RL)

$$\Delta G_{\text{bind}} = G(\text{RL}) - G(\text{R}) - G(\text{L}) \quad (1),$$

where all the reactants are assumed to be in aqueous solution. The free energy of each of the reactants is estimated as a sum of four terms:

$$G = \langle E_{\text{MM}} \rangle + \langle G_{\text{Solv}} \rangle + \langle G_{\text{np}} \rangle - T\langle S_{\text{MM}} \rangle \quad (2),$$

where G_{Solv} is the polar solvation energy of the molecule, estimated by the solution of the Poisson–Boltzmann (PB) equation [7], G_{np} is the non-polar solvation energy, estimated form the solvent-accessible surface area of the molecule [8], T is the temperature, S_{MM} is the entropy of the molecule, estimated from a normal-mode analysis of harmonic frequencies calculated at the molecular mechanics (MM) level [Error: Reference source not found], and E_{MM} is the MM energy of the molecule, i.e. the sum of the internal energy of the molecule (i.e. bonded terms, E_{int}) and the electrostatics (E_{es}) and van der Waals interactions (E_{vdW}):

$$E_{\text{MM}} = E_{\text{int}} + E_{\text{es}} + E_{\text{vdW}} \quad (3)$$

All the terms in Eqn. (2) are averages of energies obtained from a number of snapshots taken from MD simulations. In order to reduce the time-consumption and to obtain stable energies, the same geometry is normally used for all three reactants (complex, ligand and receptor), i.e. only the RL complex is explicitly simulated by MD [9]. Thereby, E_{int} cancels in the calculation of ΔG_{bind}. The MM/PBSA method has successfully been applied to many systems [Error: Reference source not found,10,11,12,13].

Despite the general success of the MM/PBSA approach, the method often has problems to discriminate between ligands with similar affinity [14,15]. The reason for this is probably that the various terms in Eqns. 2 and 3 typically show quite large variation in the various MD snapshots. For example, the standard deviation for ΔG_{bind} was 47–62 kJ/mol in a recent investigation considering the binding of seven biotin analogues to avidin [Error: Reference source not found] (of course the standard deviation of the mean value is a factor of $\sqrt{n}$ lower, where n is the number of snapshots). This variation was strongly dominated by the entropy term, whereas the other terms gave standard deviations of 22 kJ/mol or lower (if we consider the sum $E_{\text{es}} + G_{\text{Solv}}$ as a single term, because they show a parallel variation [Error: Reference source not found]). This indicates that the precision of the MM/PBSA method is strongly

limited by the entropy term.

In this paper, we suggest and validate a new method to estimate the vibrational entropy in the MM/PBSA method. We show that it performs well and gives a standard deviation of only ~15 kJ/mol for the same biotin–avidin system, i.e. similar to the other terms in MM/PBSA.

Methods

Proteins

Our calculations with factor Xa have not been published before and are therefore described here in some details. The calculations were based on the crystal structure of human factor Xa in complex with 1-(3-carbamimidoyl-benzyl)-1*H*-indole-2-carboxylic acid 3-carbamimidoyl (inhibitor number 125) [16]. Seven additional similar inhibitors were selected from the same study to give a large span in binding affinities (inhibitors number 39, 47, 53, 57, 63, 103, and 127). Inhibitors 47, 53, and 125 have a double positive charge, whereas the other ones have a single positive charge. The ligands were described with the general Amber force field (GAFF) [17], whereas we used the Amber 1999 force field for the protein [18,19]. Charges for the ligand were calculated with the same method as for the Amber 1999 force field, i.e. with a RESP (restrained electrostatic potential fit) calculation [20], based on electrostatic potentials calculated with the Hartree–Fock/6-31G* method in points sampled with the Merz–Kollman scheme [21]. Prior to the RESP calculations, the geometry of the ligands was optimized at the B3LYP/6-31G** level.

Protons were added to factor Xa, assuming the standard protonation states for all Asp, Glu, Lys, Arg, Tyr, and Cys residues (this was checked with the PROPKA software [22]). For the His residues, the protonation was decided based on PROPKA calculations and from a detailed study of the hydrogen-bond structure, the surroundings, and the solvent exposure of each residue. Thus, residues 57 and 83 were protonated on the $N^{\delta 1}$ atom, residues 91, 145, and 199 on the $N^{\varepsilon 2}$ atom, and residue 13 on both atoms. This gave a total charge of +3 or +4, depending on the ligand charge. Then, the protein was solvated in an octahedral box of TIP3P water molecules [23], extending at least 8 Å from the protein (~7100 water molecules and ~25 800 atoms in total).

The complex was first optimised by 1000 steps of minimisation, keeping all atoms, except hydrogen atoms and water molecules, restrained to their crystal position with a force constant of 418 kJ/mol/Å^2. This was followed by a 20 ps MD equilibration with a constant pressure (isotropic pressure scaling with a force constant of 1 ps) and the restraining force constant reduced to 214 kJ/mol/Å^2, a 50 ps equilibration with constant pressure and no restraints, a 200 ps MD equilibration with constant volume, and finally a 2000 ps MD simulation, in which coordinates were saved every 10 ps for a total of 200 snapshots.

In all simulations, the SHAKE algorithm [24] was used to constrain all bond length and the simulations were run by the Amber 9 sander module [25]. The temperature was kept constant at 300 K using the Berendsen weak-coupling algorithm [26] with a time constant of 1 ps. The electrostatics were treated with the particle-mesh Ewald method [27] with a fourth-order B-spline interpolation and a tolerance of 10^{-5}. The MD time step was 2 fs and the non-bonded cut-off was 8 Å. The non-bonded pair list was updated every 50 fs.

The calculations with avidin and cytochrome P450 were set up in a similar way, as has been previously described [Error: Reference source not found,28,29]. To be compatible with our previous calculations, the SHAKE algorithm was not used, and consequently a smaller time step (0.5 fs) had to be used for the simulations of the binding of flunitrazepam and progesterone to cytochrome P450, [Error: Reference source not found]. Also, only 20 snapshots were sampled.

For the binding of seven biotin analogues to avidin, three different simulation protocols

were used [Error: Reference source not found], of which one was similar to the one used for factor Xa and cytochrome P450. It employed either the Amber 1994 [Error: Reference source not found] or 2003 [30] force fields, and those simulations are called 94oh and 03oh in the following [Error: Reference source not found]. The other two approaches employed either a spherical solvated simulated system (43.4 Å radius) with a reaction-field correction for long-range interactions (03sr) or a small, truncated spherical system (20 Å), without any long-range corrections (03k). In all these simulations, only 20 snapshots were sampled. Details of those calculations have been published before [Error: Reference source not found].

In addition, we considered 11 calculations with various combinations (in the MD simulations and the MM/PBSA energy calculations) of the Amber 1994 force field and four different charge sets obtained from direct quantum mechanical (QM) calculations of the charges for all residues in the protein and for 20 snapshots. The four charge sets were the QM charges from a single snapshot (QM1), the average charges over all snapshots (Aver), consensus charges from a single (Cons1) or all (Cons) snapshots [Error: Reference source not found]. The consensus charges were obtained by averaging over all residues and atoms of the same type (e.g. over the H^{γ} atom in all serine residues). Thus, there are 54788 distinct charges in the QM1 set (one set for each atom and biotin analogue simulation), 7916 in the Aver set (one for each atom), 2924 in the Cons1 set (one for each atom type and biotin analogue simulation), and 596 in Cons (one for each atom type, i.e. similar to the Amber charges). Simulations with the Amber force field sampled 20 snapshots, whereas those with the other four charge sets sampled 152 snapshots. All simulations were performed in the same way as the 94oh simulation. The results with the conventional MM/PBSA method are identical to those published before [Error: Reference source not found,Error: Reference source not found,Error: Reference source not found].

Conventional MM/PBSA calculations

The conventional MM/PBSA calculations were automatically performed with mm_pbsa module of Amber [Error: Reference source not found]. ΔG_{bind} for all the protein–ligand complexes were calculated according to Eqns. (1–3) for all snapshots in the same way as previously described [Error: Reference source not found,Error: Reference source not found,Error: Reference source not found]. The electrostatic and van der Waals energies were calculated by the sander module. The polar solvation energy was calculated with the finite-difference PB equation solver DelPhi II [31]. Parse radii [32] were employed for all atoms. In some cases, the solvation energy was also calculated by the default Generalised Born method in Amber, viz. GB^{OBC} with α, β, and γ set to 1.0, 0.8, and 4.85, respectively [33].

The entropy was estimated by a normal-mode analysis of the harmonic vibrational frequencies, calculated using the Amber nmode module. Only residues with any atom within 8 Å of the ligand in the last snapshot were included in these calculations [Error: Reference source not found]. The truncated systems were minimised using a distance-dependent dielectric constant of $\varepsilon = 4r$ and the entropies were then calculated based on standard statistical mechanics expressions [Error: Reference source not found,34].

The non-polar solvation energy was calculated from the solvent-accessible surface area (SASA), obtained with the Amber molsurf module using a probe radius of 1.4 Å. G_{np} was obtained from the SASA according to:

$$G_{np} = \gamma \, SASA + \beta \qquad (5),$$

with $\gamma = 0.0227$ kJ/mol/Å^2 and $\beta = 3.85$ kJ/mol [Error: Reference source not found].

New entropy approach

The new approach to calculate the vibrational entropy (the translational and rotational

entropy terms are not altered) differs from the conventional one in two aspects. First, a buffer region was added to the active region. It consisted of all protein residues situated between 8 and 12 Å of the ligand (the active region consisted of all residues within 8 Å of the ligand). All water molecules within 12 Å of the ligand belonged to the buffer region, even if they were closer than 8 Å (they are not considered with the conventional method). This region was kept fixed during the minimization of the active region, and the geometrical Hessian calculation and thereby also the entropy calculation were restricted to the active region. The buffer region was present only for the free receptor and for the complex – the entropy for the free ligand was estimated for the isolated ligand alone, as in the conventional method.

The second major difference between the new and conventional approach is that in the former, a dielectric constant of 1 was used (in contrast to the distance-dependent dielectric constant of 4*r* used in the conventional approach).

In practical terms, the new method was accomplished by adding a belly list (i.e. a list of atoms to keep fixed in the minimisation) to the mm_pbsa_createinput.pm script. Moreover, the makecrd routine of mm_pbsa was replaced by a program that picked the right number of water molecules closest to the ligand in each snapshot (water molecules close to the ligand are replaced by other water molecules all the time during the MD simulation). Finally, a simple change had to be added to the source code of the nmode module, to ensure that the thermal analysis was run also when some atoms are kept fixed. Details of the procedure and modifications of the Amber source code can be found in http://www.teokem.lu.se/~ulf/Methods/mm_pbsa.html.

Result and Discussion

A new entropy method with a fixed buffer region

In the MM/PBSA approach [Error: Reference source not found], entropies are calculated by classical statistical thermodynamical expressions [Error: Reference source not found], employing a normal-mode analysis of the vibration frequencies (in the same way as frequently is done in quantum mechanics [35]). Owing to the explicit matrix diagonalisation, this has a high demand on CPU time and memory and therefore the entire protein cannot be considered. Instead, only residues within ~8 Å of the ligand are usually included in the calculations [Error: Reference source not found]. To calculate the harmonic vibrational frequencies, the truncated system needs to be minimised (vibrational frequencies are only defined with respect to a minimum on the potential energy surface) and to mimic the removed surroundings, a distance-dependent dielectric constant ($\varepsilon = 4r$) is typically used. It seems likely that the large variation of the conventional entropy term is caused by this minimisation, during which the geometry may change extensively.

A simple solution to this problem is to introduce a buffer region that is kept fixed during the minimisation, thereby ensuring that the minimised structure does not differ significantly from the original MD snapshot. Of course, the buffer region is not explicitly included in the calculation of the entropies. Thereby, we can also avoid the questionable use of a distance-dependent dielectric constant in the electrostatic calculations.

We tested eight different variants of such an approach (in all cases, the active region is minimised and the entropy is calculated only for the active region, not for any atom in the buffer region):

A. Skip the buffer completely (this is the conventional MM/PBSA approach).
B. Use the conventional approach with a larger active region (all atoms within 12 Å of the ligand).
C. Minimise also the buffer region.

D. Minimise only water molecules in the buffer region.
E. Minimise only the protein atoms in the buffer.
F. Freeze all atoms in the buffer region.
G. As in F, but still with a distance-dependent dielectric constant.
H. As in F, but with a larger buffer region (atoms between 8 and 16 Å of the ligand).

The results of these eight approaches to calculate the binding entropy of biotin to avidin are shown in Table 1. It can be seen that all approaches including the buffer region (C–H) give a strongly decreased standard deviation (15–24 kJ/mol compared to 56 kJ/mol for the conventional method A). Increasing the size of the active region in the conventional method (variant B), as has sometimes been done before [36], increased the standard deviation even more to 68 kJ/mol. Different treatments of the buffer region (variants C–F) give similar results, and the simplest approach with a fixed buffer region (variant F) gives the lowest standard deviation and also has the lowest computational cost. It can be seen that the improvement comes entirely from the fixed buffer region and not from the shift of the dielectric from a distance-dependent function (variant G) to unity constant (variant F). Finally, we note that a larger buffer region (variant H) gives a similar result to a larger computational cost. Therefore, we have selected variant F for further testing.

Test calculations on avidin

We have tested the new approach for 18 different cases of the binding of seven biotin analogues to avidin, using snapshots obtained with different methods and force fields [Error: Reference source not found,Error: Reference source not found], energy calculations with various force fields, and various selections of the active and buffer regions. The results are collected in Table 2. The standard deviation in these 18*7=126 calculations is 4–23 kJ/mol, with an average of 15 kJ/mol. With the conventional approach the standard deviation is 29–61 kJ/mol with an average of 48 kJ/mol. Thus, our method reduces the standard deviation by a factor of 3 and reduces the standard deviation to a level where it no longer limits the precision of the MM/PBSA method (the $E_{es} + G_{Solv}$ sum typically has a standard deviation of 14–26 kJ/mol, when calculated with PB and 11–21 kJ/mol when calculated by GB [Error: Reference source not found]).

There is a good correlation between entropies estimated by the two methods: The correlation coefficient (r) between the two sets of entropy estimates for the seven biotin analogues is 0.50–0.96, and the correlation coefficient between the average values is 0.98 (Figure 1), which shows that the methods do not produce random noise, but actually estimate a significant variation among the seven biotin analogues. However, our new approach tends to give slightly more negative entropies than the conventional one, on the average by 12 kJ/mol for all 126 calculations, which is fairly constant for the seven biotin analogues (cf. Figure 1).

This difference is of course somewhat alarming, but it is not too unexpected, considering the quite severe approximations in the conventional approach, viz. that the system is freely minimised and a distance-dependent dielectric constant is employed. Intuitively, our new approach should be a strong improvement by keeping the structure as close to the sampled structure as possible in the entropy calculation and avoiding the use of a distance-dependent dielectric constant only in the entropy calculation. To check the quality of the two entropy estimates, we have employed them both in full MM/PBSA predictions of the binding affinity for the seven biotin analogues. Interestingly, it turns out that the conventional method gives slightly better estimates of the absolute binding affinity, with mean absolute deviations (MADs) of 12–24 kJ/mol (average 17 kJ/mol), compared to 14–25 kJ/mol (average 19 kJ/mol) for the new method. However, the predictions of the conventional method are not improved if we allow for a systematic error (i.e. after linear regression, requiring a slope of unity, MAD* in Table 2) – the average MAD* is still 17 kJ/mol. On the other hand, the

predictions of the new method are significantly improved to an average MAD* of 15 kJ/mol (6–21 kJ/mol). Considering that the Sackur–Tetrode method employed to estimate the change in the translational entropy in (both versions of) MM/PBSA is normally considered to overestimate the reduction in entropy during ligand binding [37] and a 3 *RT* term is missing [38], there is no reason to expect MM/PBSA to give accurate *absolute* binding affinities (as has also been observed for several other proteins [Error: Reference source not found,39]). Therefore, our results indicate that the new method actually both increases the precision and accuracy of the entropy term in MM/PBSA.

In addition, we observe that the new method is stable to variations in the set-up: For example the MAD between calculations using a fixed set of active and buffer regions (New1 in Table 2) or with such regions set up specifically for the last snapshot of each calculation (New2–New4 in Table 2; the two approaches typically differ by 2–5 amino acids in the active region and 4–8 amino acids in the buffer region) is 6 kJ/mol and calculations on different biotin molecules in the tetramer differ by a similar amount. Moreover, our new method reduces the difference between entropies calculated with different charges sets: The MAD of all calculations based on the same MD snapshots, but different charges in the MM/PBSA energy calculations is reduced from 9 to 4 kJ/mol. This makes the result more stable.

Finally, we note that the average entropies in Figure 1 clearly show that an estimate of the entropy from the number of rotable bonds in the ligand (as is assumed in many scoring functions [Error: Reference source not found,40,41,42,43,44]) will not work, at least not for this set of biotin analogues: The number of rotable bonds in the seven analogues are 5, 5, 6, 7, 6, 6, and 0, respectively, and there is no correlation between these values and the average entropy values, besides that the last one gives the lowest entropy.

Test on factor Xa and cytochrome P450

Finally, we have also tested the new method for two other systems, viz. the binding of eight 3-amidinobenzyl-1*H*-indole-2-carboxamide inhibitors to factor Xa [Error: Reference source not found] and the reactivity of various sites on flunitrazepam and progesterone to cytochrome P450 2C9 and 3A4. For the former case, the standard deviation of the entropy term is reduced by a factor of 2.5–3.6 (from 49–57 to 16–20 kJ/mol), as can be seen in Table 3. The estimated entropies are not significantly altered (they change by –10 to +11 kJ/mol with a signed average of –2 kJ/mol. However, the accuracy of the MM/PBSA predictions of the binding energy is slightly worse with the new entropy method (the MAD from the experimental data [Error: Reference source not found] is 23 kJ/mol compared to 19 kJ/mol), but both predictions are strongly improved with a GB solvation model (MADs of 9 and 11 kJ/mol). The reason for this will be investigated elsewhere. The average standard deviation of the binding affinity is reduced from 58 to 33 kJ/mol (26 kJ/mol with GB), showing that the precision of the method is no longer limited by the entropy.

The results for cytochrome P450 are similar: From Table 4, it can be seen that the standard deviation of the entropy term is reduced by a factor of 2.3–4.1 (from 39–66 to 12–27 kJ/mol). Again, the estimated entropies are not significantly altered (they change by –13 to +13 kJ/mol with a signed average of –2 kJ/mol). In particular, the range of the entropies is reduced from 39 to 23 kJ/mol. No correlation to experiments can be performed, because experimental data is missing for several sites (they are non-reactive).

Conclusions

We have suggested an improved method to calculate the vibrational entropy term in the MM/PBSA approach [Error: Reference source not found]. It employs a buffer region that is kept fixed during the minimisation of the active region. Thereby, we ensure that the active region remains close to the starting geometry. Moreover, we avoid the use of a distance-

dependent dielectric constant, which is inconsistent with the rest of the MM/PBSA method. Entropies have been calculated before with atoms outside 6 Å of the ligands kept fixed, but no details of the method were given, it was not compared to the conventional approach, and the resulting entropies were not evaluated [Error: Reference source not found].

We have compared the new and conventional approaches for three different protein–ligand systems, viz. the binding of seven biotin analogues to avidin, using different simulation methods, force fields, and charges [Error: Reference source not found,Error: Reference source not found,Error: Reference source not found], the binding of a 3-amidinobenzyl-1*H*-indole-2-carboxamide inhibitor to human factor Xa [Error: Reference source not found], and prediction of the reactivity of various sites of flunitrazepam and progesterone to human cytochromes P450 2C9 and 3A4 [Error: Reference source not found]. In all cases, we obtain a reduction of the standard deviation of the vibrational entropy estimates from the various snapshots by a factor of 2–4. Such a reduction ensures that the entropy term no longer limits the precision of the MM/PBSA method (instead, the sum of the electrostatics and the solvation energy is limiting).

For avidin, for which we have a large statistical material of 33 series of calculations, the magnitude of the entropy calculated with the new method slightly increases (12 kJ/mol) compared to the corresponding predictions based on the conventional method. This leads to slightly improved relative (but not absolute) MM/PBSA estimates of the binding affinities. For the other two test cases, for which only one series is studied, no such trends are observed. In fact, the new method leads to slightly worse binding affinity estimates for factor Xa, but this may be accidental (this also happen for 4 of 15 comparisons for avidin). Moreover, MM/PBSA involves several other severe approximations, so it cannot be expected that the accuracy of the total binding affinity estimates always are improved even if the individual terms are improved. The same was observed in our previous study of the biotin–avidin complex, in which no significant improvement was seen by using a polarisable force field or by using charges calculated by quantum-mechanical methods on the actual geometry for all amino acids in the protein [13]

Finally, we also note that the new method is more stable: It gives more similar results for calculations based on the same MD snapshots but employing different charges in the energy calculations. Moreover, the conventional version occasionally crashes because there are too large geometrical changes in the system (it extends outside the virtual simulation box). In conclusion, we expect that this new approach for the entropies will increase the usefulness of MM/PBSA for ligand-binding studies.

Acknowledgements

This investigation has been supported by grants from the Swedish research council and the Villum Kann Rasmussen foundation (Denmark), and by computer resources of Lunarc at Lund University.

References

Table 1. Results for the five variants (A–E) of the new MM/PBSA entropy (see the text). The calculated entropy ($T\Delta S$) and the corresponding standard deviation (Stdev) are listed (both in kJ/mol).

Variant	$T\Delta S$	Stdev
A	89.5	55.5
B	78.7	67.6
C	98.2	23.0
D	101.6	24.4
E	102.2	20.1
F	97.5	17.3
G	96.0	15.8
H	99.7	15.4

Table 2. Results for the biotin–avidin calculations, showing the estimated entropy ($T\Delta S$) and the corresponding standard deviation for the seven biotin analogues, the mean absolute difference of the estimated binding affinities to experimental data [Error: Reference source not found], without (MAD) or with (MAD*) the signed average value subtracted (all entries in kJ/mol). MD and Energy indicate the force fields and charge sets used to generate the MD snapshots and in the MM/PBSA energy calculations, respectively. The number of the entropy method indicates if the same definition of the active and buffer regions has been used (same number) or not.

MD	Energy	Entropy	$T\Delta S$							Standard deviation								
		method	1	2	3	4	5	6	7	1	2	3	4	5	6	7	MAD	MAD*
94oh	94	Old	-89.4	-91.5	-93.3	-91.5	-66.3	-61.2	-58.3	55.4	50.3	60.5	48.1	34.2	55.3	38.7	13.5	14.0
		New	-98.5	-105.2	-108.3	-109.7	-87.3	-76.4	-67.6	14.3	20.5	21.8	15.3	13.5	16.4	6.0	18.0	15.6
94sr	94	Old	-77.4	-85.3	-87.5	-75.0	-78.9	-88.2	-60.9	48.0	35.9	60.2	41.5	44.9	44.6	37.9	19.0	18.8
		New	-96.9	-109.8	-104.7	-93.3	-85.9	-80.0	-63.5	17.8	12.1	13.8	17.9	14.9	8.4	7.8	19.1	12.4
94k	94	Old	-72.2	-97.4	-89.7	-78.2	-73.2	-67.2	-62.2	41.8	47.9	53.5	52.7	53.3	50.3	40.7	16.3	15.9
		New	-114.0	-107.5	-106.1	-95.0	-85.8	-84.3	-67.8	18.1	20.0	16.8	16.8	14.2	13.6	8.1	21.6	11.3
03oh	03	Old	-81.4	-96.2	-69.7	-81.5	-66.8	-65.5	-28.0	45.6	61.1	55.9	53.5	51.8	49.8	57.1	15.9	16.1
		New	-102.1	-106.8	-103.6	-103.6	-83.0	-77.9	-59.9	15.0	15.6	22.3	21.2	15.7	14.0	10.6	24.5	12.2
94oh*	94	Old1	-89.4	-91.5	-93.3	-91.5	-66.3	-61.2	-58.3	55.4	50.3	60.5	48.1	34.2	55.3	38.7	13.5	14.0
		New1	-96.9	-96.5	-107.2	-109.9	-85.6	-79.8	-68.0	16.7	17.5	18.0	19.7	12.8	12.6	4.6	19.7	17.8
	Aver	Old1	-105.4	-90.5	-74.1	-110.0	-75.4	-78.2	-67.2	49.5	56.6	44.5	36.0	45.6	55.6	29.1	23.6	23.2
		New1	-94.5	-95.9	-108.8	-101.7	-87.8	-77.7	-68.8	13.4	13.3	17.4	13.6	11.5	10.7	5.6	20.2	19.0
	Cons	Old1	-95.3	-105.8	-104.4	-87.9	-64.4	-90.5	-57.1	45.9	43.2	58.4	45.7	33.8	54.6	34.4	15.7	15.2
		New1	-99.0	-102.5	-102.9	-103.6	-85.9	-81.2	-69.7	18.1	22.3	17.2	14.4	10.3	14.1	4.2	21.9	20.6
	Cons1	Old1	-82.8	-106.4	-95.8	-88.3	-86.1	-54.8	-65.2	55.0	57.3	43.1	49.4	29.2	58.1	39.4	18.5	18.5
		New1	-96.9	-101.5	-107.8	-100.6	-79.5	-80.3	-69.0	20.1	20.3	16.7	16.9	13.4	17.5	7.6	20.0	18.8
	QM1	Old1	-77.8	-104.5	-84.5	-95.4	-88.4	-70.3	-58.3	47.2	44.7	30.9	56.4	44.2	48.5	38.7	24.3	23.5
		New1	-100.6	-99.8	-101.5	-100.7	-84.7	-76.4	-69.2	22.0	15.3	21.8	12.5	15.7	14.3	6.6	21.9	20.0
Aver	Aver	Old1	-90.4	-72.5	-70.1	-54.1	-68.8	-61.5	-51.7	46.5	51.1	51.4	55.6	54.8	53.6	51.3	12.1	11.5
		New2	-107.0	-102.8	-103.0	-91.6	-81.5	-80.1	-62.0	17.2	16.8	17.2	16.7	13.6	12.3	11.6	18.7	6.4
	94	Old1	-90.0	-96.3	-68.4	-72.0	-78.3	-67.4	-52.6	47.2	47.9	49.9	56.1	46.9	54.2	45.6	15.8	15.5
		New2	-111.4	-101.0	-97.9	-92.0	-79.3	-79.3	-61.9	13.6	17.0	16.9	16.4	14.0	12.7	11.8	14.4	11.2
		New1	-114.1	-106.1	-100.6	-96.8	-79.4	-86.2	-66.2	18.0	20.6	18.3	15.9	13.8	16.2	9.8	17.0	14.8
Cons	Cons	Old1	-88.2	-95.4	-88.7	-89.1	-81.2	-79.3	-60.1	44.4	46.2	48.0	48.0	48.2	49.7	45.6	20.6	20.6
		New3	-104.4	-108.4	-101.5	-97.8	-81.2	-77.3	-65.6	15.9	16.1	15.9	14.7	13.6	12.8	8.3	19.9	19.9
	94	Old1	-90.3	-93.8	-93.6	-83.3	-78.9	-72.3	-52.8	49.6	49.1	50.7	50.2	40.6	47.0	46.9	15.9	16.3
		New3	-102.1	-106.8	-100.0	-95.9	-76.3	-78.5	-66.3	14.6	17.6	16.9	15.1	14.6	13.1	8.0	17.5	14.2
		New1	-119.7	-113.5	-107.5	-92.3	-82.5	-80.8	-65.4	17.4	20.3	18.2	16.9	17.3	13.7	10.5	18.6	10.2
Cons1	Cons1	Old1	-91.0	-83.3	-94.8	-88.7	-82.8	-73.2	-61.5	47.4	51.6	46.9	50.9	42.6	48.2	46.7	17.9	17.8
		New4	-103.1	-110.9	-99.6	-90.0	-83.2	-78.5	-75.9	16.7	16.0	23.3	16.2	14.6	12.6	14.4	15.7	13.7
	94	Old1	-93.1	-82.6	-97.9	-91.6	-78.0	-44.6	-57.1	44.4	52.6	46.6	51.6	46.3	48.0	45.6	16.6	16.2
		New4	-103.1	-104.3	-102.7	-92.1	-78.9	-80.6	-78.2	15.6	15.8	21.9	15.3	13.2	14.2	15.2	17.3	14.0
		New1	-106.3	-106.2	-96.1	-101.8	-80.1	-84.1	-67.8	16.7	17.4	16.6	16.9	13.8	14.3	8.1	19.0	14.7

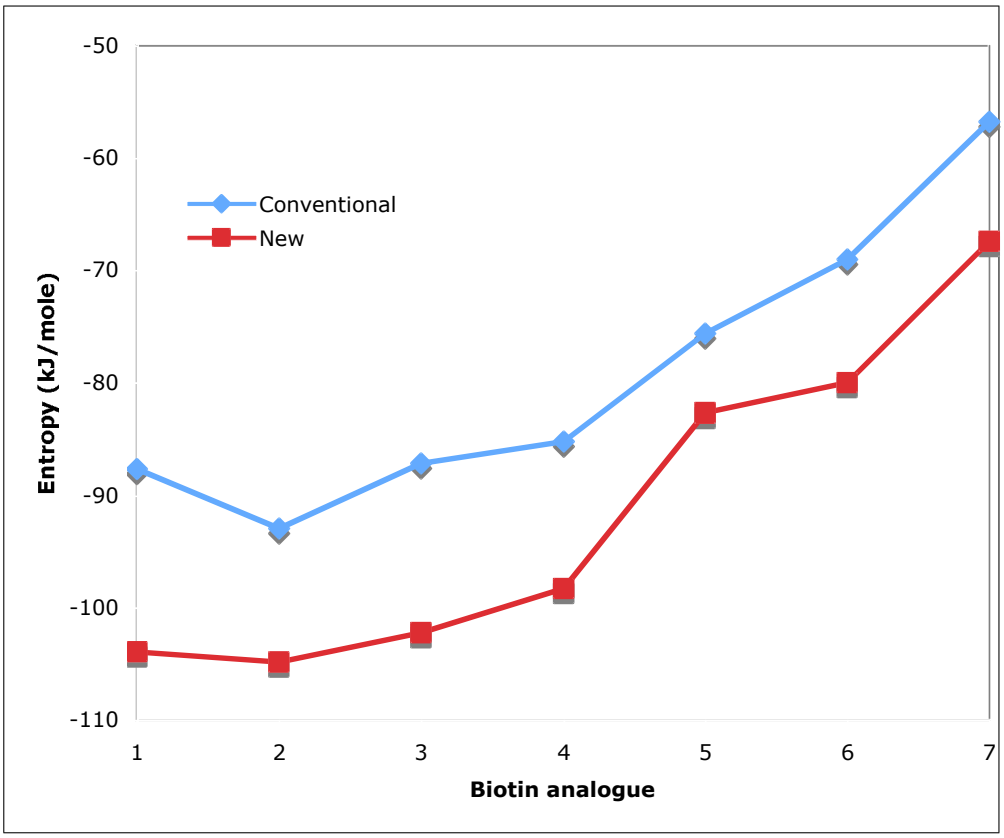
Table 3. Results for the factor Xa calculations, showing the estimated entropy ($T\Delta S$) and the corresponding standard deviation (Stdev). All the results are in kJ/mol.

Quantity	Method	39	47	53	57	63	103	125	127
$T\Delta S$	Old	-113.8	-113.9	-106.3	-120.6	-123.5	-110.9	-101.7	-90.5
	New	-120.0	-119.2	-113.2	-115.6	-112.5	-111.3	-112.0	-91.4
Stdev	Old	51.4	57.1	51.9	54.0	53.3	48.7	53.5	57.2
	New	20.2	19.1	17.9	16.3	16.8	16.4	16.2	15.7

Table 4. Results for the cytochrome P450 calculations, showing the estimated entropy ($T\Delta S$) and the corresponding standard deviation (Stdev). All the results are in kJ/mol.

		Flunitrazepam				Progesterone			
		2C9		3A4		2C9		3A4	
Quantity	Method	H1	H3	H1	H3	16a	16b	16a	16b
$T\Delta S$	Old	-57.7	-96.9	-72.7	-90.6	-80.5	-71.9	-90.7	-86.2
	New	-70.4	-83.6	-80.2	-81.5	-93.8	-84.0	-79.0	-87.7
Stdev	Old	57.1	39.1	49.2	55.7	63.2	50.9	49.0	66.2
	New	15.4	17.0	11.9	15.9	23.6	13.6	19.2	26.8

Figure 1. The relation between entropic contributions to the binding energy of seven biotin analogues to avidin, estimated with the conventional and new methods. Averages over 15 (conventional) or 18 (new) calculations.



1 Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew. Chem. Int. Ed. 2002, 41, 2644-2676.

2 Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical thermodynamic basis for computation of binding affinities: A critical review. Biophys. J. 1997, 72, 1047-1069.

3 Beveridge, D. L.; Dicapua, F. M. Free-energy via molecular simulation - applications to chemical and biomolecular systems. Annual Review of Biophysics and Biophysical Chemistry 1989, 18, 431-492.

4 Miyamoto, S.; Kollman, P. A. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. Proteins, Struct. Funct. Genet. 1993, 16, 226-245.

5 T. Hansson, J. Marelius, J. Åqvist, Ligand binding affinity prediction by linear interaction energy methods. J. Comput.-Aided Mol. Design 1998, 12, 27-35.

6 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A . Case, T. E. Cheatham, Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. Acc. Chem. Res. 2000, 33(12), 889-897.

7 Gilson, M. K; Honig, B. Calculations of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. Proteins, Struct. Funct. Gen. 1998, 4, 7-18.

8 Hermann, R. B. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. J. Phys. Chem. 1972, 76, 2754-2759.

9 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.

10 B. Kuhn, P. A. Kollman, 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.

11 S. Huo, J. Wang, P. Cieplak, P. A. Kollman, I. D. Kuntz, Molecular Dynamics and Free Energy Analyses of cathepsin D-Inhibitor Interactions: Insight into Structure-Based Ligand Design, J. Med. Chem. 2002, 45, 1412-1419.

12 Brown, S. P; Muchmore, S. W. High-throughput calculation of protein-ligand binding affinities: Modification and adaptation of the MM-PBSA protocol to enterprise grid computing. J. Chem. Inf. Model. 2006, 46, 999-1005.

13 A. Weis, K. Katebzadeh, P. Söderhjelm, I. Nilsson, U. Ryde (2006) “Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field”, J. Med. Chem., 49, 6596-6606.

14 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.

15 Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. J. Med. Chem. 2005, 48, 4040-4048.

16 H. Matter, E. Defossa, U. Heinelt, P.-M. Blohm, D. Schneider, A. üller, S. Herok, H. Schreuder, A. Liesum, V. Brachvogel, P. Lönze, A. Walser, F. Al-Obeidi, P. Wildgoose, J. Med. Chem. 45 (2002) 2749-2769.

17 Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. J. Comput. Chem. 2004, 25, 1157-1174.

18 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 1995, 117, 5179-5197.

19 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.

20 C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints For Deriving Atomic Charges - the RESP Model, J. Phys. Chem. 1993, 97, 10269-10280.
21 B. H. Besler, K. M. Merz, P. A. Kollman, Atomic Charges Derived from semiempirical Methods, J. Comput. Chem. 1990, 11, 431-439.
22 H Li, A. D. Robertson, J. H. Jensen "Very Fast Empirical Prediction and Interpretation of Protein pKa Values" *Proteins*, 2005, 61, 704-721.
23 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein.; Comparison of Simple Potential Functions for Simulating Liquid Water, J. Chem. Phys. 1983, 79, 926-935.
24 J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, 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.
25 D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman (2006), AMBER 9, University of California, San Francisco.
26 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 1984, 81, 3684-3690.
27 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.
28 P. Söderhjelm, U. Ryde. Conformational dependence of charges in protein simulations. J. Comput. Chem., in press (2008).
29 P. Rydberg, S. M. Hansen, J: Kongsted, P.-O. Norrby, L. Olsen, U. Ryde, J. Chem. Theory Comput., (2008) 4, 673-681.
30 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee.; A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins, J. Comput. Chem. 2003, 24, 1999-2012.
31 W. Rocchia, E. Alexov, B. Honig.; Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions, J. Phys. Chem. B 2001, 105:6507-6514.
32 D. Sitkoff, K. A. Sharp, B. Honig, Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models, J. Phys. Chem. 1994, 98, 1978-1988.
33 A. Onufriev, D. Bashford, D. A. Case, Exploring protein native states and large-scale conformational changes with a modified generalized Born model, Proteins 2004, 55, 383-394.
34 Jensen, F. Introduction to Computational Chemistry 1999, John Wiley & Sons.
35 Siegbahn, P. E. M.; Blomberg, M. R. A. Chem. Rev. 2000, 100, 421-437.
36 Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. J. Am. Chem. Soc. 2001, 123, 5221-5230.
37 Amzel L. M. Proteins Struct Funct Gen 1997, 28, 144-149.
38 Gohlke, R.; Case, D. A. J. Comput. Chem. 2004, 25, 238-250.
39 A. Suenaga, N. Takada, M. Hatakeyama, M. Ichikawa, X. Yu, K. Tomii, N. Okimoto, N. Futatsugi, T. Narumi, M. Shirouzu, S. Yokoyama, A. Konagaya, M. Taiji, J. Biol. Chem. 280 (2005) 1321-1326.
40 Searle, M. S.; Williams, D. H. J. Am. Chem. Soc. 1992, 114, 10690.
41 Böhm, H.-J. J. Comput.-Aided Mol. Design 1994, 8, 243-256.
42 Lazardis, T.; Masunov, A.; Gandolfo, F. Proteins, Struct. Funct. Gen. 2002, 47, 194-208.
43 Giordanetto, F.; Cotesta, S.; Catana, C.; Trosset, J.-Y.; Vulpetti, A.; Stouten, P. F. W.; Kroemer, R. T. J. Chem. Inf. Comput. Sci. 2004, 44, 882-893.
44 Raha, K.; Merz, K. M. J. Am. Chem. Soc. 2004, 126, 1020-1021.