LUND UNIVERSITY

Ligand affinities estimated by quantum chemical calculations

Söderhjelm, Pär; Kongsted, Jacob; Ryde, Ulf

Link to publication

Total number of authors:
3

Ligand affinities estimated by

quantum chemical calculations

Pär Söderhjelm, Jacob Kongsted, and Ulf Ryde *

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

*Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se,

Abstract

We present the first fully quantum chemical estimates of ligand-binding affinities, performed at a reasonable level of theory (MP2/cc-pVTZ) and at the same time including all physically important effects, such as solvation, entropy, and sampling. We have studied the binding of seven biotin analogues to the avidin tetramer. The calculations have been performed by the recently developed PMISP approach (polarisable multipole interactions with supermolecular pairs), which treats electrostatic interactions by multipoles up to quadrupoles, induction by anisotropic polarisabilities, and non-classical interactions (dispersion, exchange repulsion, etc.) by explicit quantum chemical calculations, using a fragmentation approach, except for long-range interactions, which are treated by standard molecular mechanics Lennard-Jones terms. In order to include effects of sampling, ten snapshots from a molecular dynamics simulation are studied for each complex. Solvation energies are estimated by a polarised continuum model (PCM), coupled to the multipole–polarisability model. Entropy effects are estimated from vibrational frequencies, calculated at the molecular mechanics level. We encounter several problems, not previously discussed, illustrating that we are first to apply such a method. For example, the PCM model is questionable for large molecules, owing to the use of a surface definition that gives numerous small cavities in a protein.

Introduction

A major goal of theoretical chemistry is to accurately predict the free energy for the binding of a ligand to a macromolecule. If such binding affinities could be accurately predicted, large parts of the drug development could be performed by computer simulations rather than by costly experiments, because essentially all drugs evoke their action by binding to a target macromolecule. Likewise, many interesting questions in biochemistry can be formulated as the differential binding affinities of a substrate, product, or transition-state to a protein or enzyme.

Consequently, numerous theoretical methods have been developed to estimate ligand affinities [1]. The most accurate ones are based on free-energy perturbation and related approaches [2]. Unfortunately, they are extremely time-consuming and the results typically converge only in the case where the difference in binding affinity of similar ligands are considered, i.e. for relative binding affinities. Therefore, many more approximate methods have been suggested. Some of them are still based on extensive sampling of the phase space, e.g. the linear interaction energy (LIE) and molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) approaches [3,4]. Other methods use a single molecular conformation and estimate the binding affinities by methods based on either physics or statistics [1].

Most of the physical methods are based on calculations with a molecular mechanics (MM) force field. Clearly, the accuracy of such methods can never be better than that of the force field. Therefore, there has recently been a great interest in developing ligand-binding methods that are based on quantum mechanics (QM), rather than on a MM force field [5]. Such methods are typically based on either semiempirical calculations [6,7] or on higher-level methods, based on fractionation approaches, e.g. the fragment molecular orbital method (FMO) [8,9] or the molecular fractionation with conjugate caps (MFCC) and related methods [10,11,12,13,14,15,16]. It is well-known that calculations of dispersion effects generally require a very high level of theory [17]. Likewise, accurate predictions of polarisation effects and dispersion require the use of a large and flexible one-electron basis set [17,18]. Only one of these previous studies [16] has been performed at a level (MP2/6-311(+)G(2d,p)), for which there is hope that dispersion and polarisation effects are treated balanced and satisfactorily.

Recently, we have developed an approach that is intended to provide accurate interaction energies between a ligand and a macromolecule at a proper level of theory [19]. It is called PMISP (polarisable multipole interactions with supermolecular pairs). It treats electrostatic interactions by multipoles up to quadrupoles, induction by anisotropic polarisabilities, and non-classical interactions (dispersion, exchange repulsion, etc.) by explicit quantum mechanical calculations, using a fragmentation approach similar to MFCC. At a given level of QM theory, it can be seen as the best possible MM force field that treats all terms, except polarisation pair-wise. It gives an accuracy of 2–5 kJ/mol for neutral and ~10 kJ/mol for charged ligands compared to full QM treatments [19]. For calculations with a full protein, much computer time can be saved if long-range interactions are treated by a QM/MM approach, i.e. by a standard Lennard-Jones term (PMISP/MM) [20]. If the boundary between the PMISP and MM systems are chosen far enough from the ligand, this approximation does not add any additional uncertainty. By this approach, we have illustrated the importance of using a proper level of theory. For example, the interaction energy between biotin and avidin differ by 150 kJ/mol if calculated at the MP2 level with the 6-31G* or aug-cc-pVTZ basis sets [20], showing that the former estimate is quite meaningless.

However, in order to provide reliable ligand-binding energies, more terms than the pure interaction energy need to be considered. In particular, the effects of the surrounding solvent, entropy, and sampling need to be taken into account [1,4]. Only a few of the previous studies take into account effects of solvation [6,7,9] and entropy [6,7], and none of them consider sampling.

In this paper, we present what seems to be the first realistic QM estimation of ligand-binding affinities at a proper level of theory and at the same time taking into account the combined effects of solvation, entropy, and sampling. We employ the PMISP/MM method at the MP2/cc-pVTZ level within the framework of the MM/PBSA method. We study the affinities of seven biotin analogues to the full avidin tetramer. This system it is well characterised by X-ray crystallography [21,22,23,24] and experimental binding free energies for a number of ligands (biotin analogues) are available [25,26,27]. Moreover, it has been investigated by several different theoretical methods [28,29,30,31,32,33].

Methods

The PMISP/MM method

The PMISP and PMISP/MM approaches have been thoroughly described before [19,20]. Therefore, we here only provide a short summary of the methods. We consider the binding of a small ligand (*L*) to a protein (*P*):

$$P + L \rightarrow PL \tag{1}$$

In the PMISP method [19], the interaction energy of this reaction is estimated by:

$$E_{\text{PMISP}}(PL) = E_{\text{el}}(PL) + E_{\text{ind}}(PL) + E_{\text{nc}}(PL) \tag{2},$$

where E_{el} and E_{ind} are the electrostatic and induction interaction energies, respectively (note that all energies in Eqns. 2–4 are interaction energies between *L* and *P*, not the absolute energies of the *PL* complex). E_{el} is calculated from a multicentre–multipole expansion up to quadrupoles, centred at all atoms and bond midpoints in the protein and the ligand. Likewise, E_{ind} is calculated from anisotropic dipole polarisabilities in the same centres in a self-consistent manner. Both these terms are obtained with the LoProp approach [34]. E_{nc} is the non-classical term, containing mainly dispersion and exchange repulsion, but also short-range corrections to the classical terms, e.g. charge penetration. It is estimated by

$$E_{nc}(PL) = \sum_{i=1}^{n} c_i \left(E_{QM}(P_i L) - E_{el}(P_i L) - E_{ind}(P_i L) \right) \tag{3},$$

where the protein has been divided into a number of fragments (P_i), using the molecular fractionation with conjugate caps (MFCC) method [35]. In this paper, each amino acid constitutes one fragment, and they are capped with $CH_3CO–$ and $–NHCH_3$ groups. The caps from neighbouring fragments are joined to form a $CH_3CONHCH_3$ conjugated cap (concap) for each peptide bond and the energies of these concaps are subtracted ($c_i = -1$ in Eqn. 3) from the energies of the capped amino-acid fragments ($c_i = 1$). This has been shown to be an excellent approximation, giving errors of only ~1 kJ/mol [19]. $E_{QM}(P_iL)$ is the counterpoise-corrected quantum mechanical (QM) interaction energy of the P_i–L pair. A similar formula is used to derive properties (multipoles and polarisabilities) for the whole protein from fragment-wise calculations [19]. E_{QM} was calculated at the MP2/cc-pVTZ level, which has been shown to provide dispersion energies similar to coupled-cluster methods with larger basis sets, owing to error cancellation [20,36]. The multipoles and polarisabilities were calculated at the B3LYP/6-31G* level, which has been shown to be a good approximation for the much more expensive MP2/cc-pVTZ properties, provided that the same properties are used in both Eqns. 2 and 3 [20].

For a large protein, only a few fragments P_i are in close contact with the ligand, so the direct use of Eqn. 2 would be very inefficient. Therefore, we can save much time without compromising the accuracy by using a QM/MM approach, PMISP/MM [20]: For a model containing residues close to the ligand (*M*), the full PMISP approach is used, whereas for more distant residues, E_{nc} is approximated by the Lennard-Jones term from a classical force field, E_{LJ}:

$$E_{\text{PMISP/MM}}(PL) = E_{\text{el}}(PL) + E_{\text{ind}}(PL) + E_{\text{nc}}(ML) + E_{\text{LJ}}(PL) - E_{\text{LJ}}(ML) \tag{4}$$

Thus, we use the same accurate multipole–polarisability model for the whole protein as in PMISP. In this work, the E_{LJ} term is taken from the Amber 1994 force field (the same terms are also used in the newer Amber 2003 and the polarisable 2002 force fields) [37,38,39]. Naturally,

the accuracy of this approximation will improve as the size of the *M* region is increased [20]. In this work, we have used all atoms within 4 Å of the ligand and added enough atoms to obtain chemically reasonable groups, such as aromatic rings or amide groups. For groups that form exceptionally strong interactions with the ligand (distances shorter than 1.7 Å), the model was extended with an extra CH_2 group, to avoid the largest errors observed previously [20] (e.g. Ser-73 was modelled by ethanol, rather than methanol). Thus, *M* consisted of 165–271 atoms, depending on the ligand (but the same *M* region was used for all snapshots with the same ligand). All PMISP calculations were performed with the Molcas 7.2 software [40], using Cholesky decomposition in combination with the local exchange algorithm [41,42,43].

Solvation calculations with the PCM method

To accurately estimate ligand-binding affinities, an accurate estimate of the change in solvation energy upon ligand binding is needed. The standard continuum solvation methods for MM/PBSA in the AMBER software [44], the Poison–Boltzmann or generalised Born methods, cannot handle a multipole expansion or polarisabilities (although there exist recent extensions that can [45,46]). Therefore, we instead decided to use the PCM method, which has recently been extended to be used with the effective fragment potential method (which also uses a polarisable force field with a multipole expansion) [47]. We used the integral equation formulation of PCM, IEFPCM [48]. Owing to the large size of the molecular systems, the PCM problem was solved using a direct inversion of the iterative subspace procedure [47], as implemented in the GAMESS software [49]. The PCM calculations, were performed at the MM level. Thus, we constructed the PMISP multipole and polarisable force field for the solute and this force field was then used in the calculation of the solvation energies.

Like all continuum-solvation approaches, PCM employs dielectric cavities defined by a set of atomic radii. For accurate predictions of solvation energies, it is mandatory to use optimised cavity parameters. Several such sets of parameters are available for PCM at various levels of theory, e.g. Hartree–Fock [50] and density functional theory (UAHF and UAKS, i.e. united-atom topological model for Hartree–Fock and Kohn–Sham theory). Since we base our predictions on B3LYP and MP2 calculations, we decided use the latter radii, which were optimised using the PBE0 functional. These radii, although not yet properly published, are available in the Gaussian-03 suite of programs [51].

Since we use the UAKS parameters at the MM level, recalibration of these parameters is strictly needed. However, we limited the recalibration to a scaling of the radii for the electrostatic component in the PCM solvation energy calculation. For the original UAKS radii, this scaling parameter is 1.2. The calibration was based on a test-set of 22 small organic molecules, listed in Table 1. These molecules were selected to represent models of the peptide backbone and all amino-acid side-chains. For these,we constructed distributed multipoles up to quadrupoles and anisotropic polarisabilities in the same way as for PMISP [19]. The multipoles and polarisabilities were calculated using the B3LYP/6-31G* method (6-31+G* for the two anions) and the solvation energies were then evaluated using the PCM approach for various values of the scaling parameter. The non-polar solvation terms (cavitation, dispersion and exchange repulsion [50]) are independent of this scaling factor and were therefore calculated only once. As will be discussed below, we encountered serious problems with the non-polar terms in the PCM model. Therefore, the final calibration of the PCM method (Table 1) employed instead the non-polar energy from the standard MM/PBSA method. A fitting to experimental data [52,53,54,55,56], gave a scaling factor of 1.12 (with the non-polar terms from PCM, the optimum scaling factor was 1.15). This decrease in the scaling factor is expected, because at the MM level, there is no charge penetration. The scaled model gave MADs of 2 and 4 kJ/mol, for the neutral molecules and all molecules, respectively. This is only slightly worse than for the UAHF parameters (1 and 3 kJ/mol), similar to the UAKS parameters (1 and 5 kJ/mol), and appreciably better than seven different Poison–Boltzmann and eleven

4

generalised Born methods [57] (3–9 and 7–18 kJ/mol). For the seven biotin analogues, this recalibrated PCM method gives a MAD of 10 kJ/mol, compared to a weighted average of 24 different continuum solvation methods [57], which again is slightly worse than for the original UAHF and UAKS methods (4 and 7 kJ/mol).

MM/PBSA

The calculations in this paper are based on the MM/PBSA approach [4]. In this method, the binding affinity (the free energy of the reaction in Eqn. 1, ΔG_{bind}) is estimated from the free energies of the three reactants,

$$\Delta G_{bind} = G(PL) - G(P) - G(L) \tag{5}$$

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

$$G = <E_{MM}> + <G_{solv}> + <G_{np}> - T<S_{MM}> \tag{6}$$

where G_{solv} is the polar solvation energy of the molecule, estimated by the solution of the Poisson–Boltzmann (PB) equation [58], G_{np} is the non-polar solvation energy, estimated form the solvent-accessible surface area (SASA) of the molecule [59], 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, 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_{bonded}), the electrostatics (E_{es}), induction energy (E_{ind}, only if a polarisable force field is used), and van der Waals interactions (E_{vdW}):

$$E_{MM} = E_{bonded} + E_{es} + E_{ind} + E_{vdW} \tag{7}$$

All the terms in Eqn. (6) 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 *PL* complex is explicitly simulated by MD [60]. Thereby, E_{bonded} cancels in the calculation of ΔG_{bind}.

In this investigation, we test if the binding-affinity predictions can be improved by replacing some of these terms with estimates using other methods. Thus, we replace the E_{MM} term by the PMISP/MM estimate of the interaction energy between the ligand and the protein. Second, we replace the $G_{solv} + G_{np}$ estimates of the solvation energies by the corresponding terms within the PCM model. It should be noted that because of the non-additivity of the induced energy, there is no unique way to separate the E_{ind} and G_{solv} terms. However, we found that the default decomposition done in GAMESS [49] gave reasonable trends and we therefore use this decomposition in the discussion. Other approaches to replace the E_{MM} term with a standard QM/MM term have been tested, both for calculations of ligand-binding affinities and for other energies [61,62,63].

Thus, only the S_{MM} term is kept from the original MM/PBSA method, but it is calculated according to our recently developed improved method [64]. In the original approach [4], the protein is truncated 8 Å from the ligand and it is then freely optimised, using a distance-dependent dielectric constant $\epsilon = 4r$. We have shown that this gives a large statistical uncertainty in the entropy estimate, which can be reduced by a factor of 2–4 if a buffer region of 4 Å is used outside the cut-off radius. This buffer region is kept fixed in the geometry optimisation and not included in the estimate of the entropy, but it ensures that the optimised system stays close to the complex structure. This also makes the use of the questionable

distance-dependent dielectric constant superfluous.

The results of the PMISP/MM/PCM/$T\Delta S$ approach is compared with the results of standard MM/PBSA calculations using the polarisable Amber 2002 force field [39]. These were performed in the same way as in our previous investigation of various force fields for the biotin–avidin system (02ohp/02 calculation; but the present calculations are based on the ligand in the fourth subunit in the tetramer, rather than the first one in the previous investigation) [33]. This means that G_{solv} is estimated by adding an extra charge close to each atom site to simulate the induced dipoles in the PB calculations. Calculations of both G_{solv} and G_{np} used Parse radii [65]. Moreover, the entropy estimate was done with our new method (this term is identical to the one used in the PMISP estimate. Unfortunately, the Amber nmode program does not work properly for a polarisable force field (in contrast to what is indicated in the manual), so the entropy calculations were performed without the polarisabilities.

Studied systems

The seven biotin analogues (BTN1–BTN7) studied in this investigation are the same as in ref. [33]. The set-up of the molecular dynamics simulations have been described before [33]. We used ten snapshots (sampled every 20 ps) for each analogue taken from this investigation, performed by the polarisable Amber 2002 force field [39] (the 02ohp simulation).

Result and Discussion

Non-polar solvation energy

First, we tried to calculate the solvation energies using the full PCM model implemented in the GAMESS program [49]. However, this gave differential solvation energies (i.e. $G_{solv}(PL) - G_{solv}(P) - G_{solv}(L)$) that were 60–140 kJ/mol more positive than the corresponding results with a PB+SASA model. Further inspection shows that the difference arises almost entirely from the non-polar part of the solvation energy: In the PB+SASA method, this term is taken from the difference in solvent-exposed surface area between the complex and the isolated protein and ligand. From the results presented in Table 2, it can be seen that the SASA non-polar energies are quite small and similar for the complex and the protein, ~470 kJ/mol (corresponding to a SASA of 20 600 Å^2, because ΔG_{np} = SASA*0.0227 – 3.85 kJ/mol, when SASA is given in Å^2 [33]). The difference is 1–4 kJ/mol, with the protein having the largest value, which indicates that only ~25% of the ligand contributes to the SASA in the complex (i.e. the ligand is mainly buried in the protein). Therefore, the net non-polar SASA effect comes mostly from the ligand. As an effect, ΔG_{np} in MM/PBSA is small and positive for all complexes, 11–21 kJ/mol and directly related to the size of the ligand.

However, in the PCM method, the non-polar solvation energy is calculated from three separate terms: the energy cost of making a cavity in the solvent (the cavitation energy), a favourable term from the dispersion interactions between the solute and the solvent, and the corresponding unfavourable term from exchange repulsion [50]. The former term is calculated from an expression that contains terms involving the radius of each atom to the power of 0–3 [66], i.e. including a term that is proportional to the the volume, whereas latter two terms are calculated by a surface-based integration method [67]. In fact, the PCM energies are almost 50 times larger than the SASA energies, ~22 200 kJ/mol. The PCM energies are dominated by the cavitation energy, which is ~28 000 kJ/mol, compared to the dispersion energy of ~−7500 kJ/mol and the exchange repulsion energy of ~2 000 kJ/mol (cf. Table 2). However, when computing the difference upon binding, the cavitation energy is mainly cancelled (the net effect is negative and 8–22 kJ/mol). This indicates that the volume term of the cavitation energy is dominating the individual energies, because the volume hardly changes during ligand binding.

On the other hand, the surface area is reduced, and this causes a positive (unfavourable) contribution from the dispersion term of 89–236 kJ/mol, only partly cancelled by the exchange repulsion (negative and 29–61 kJ/mol) and by the small cavitation energy. The net ΔG_{np} is therefore 52–155 kJ/mol, i.e. it has the opposite sign and is larger in magnitude compared to the SASA non-polar solvation energy. It is notable that the two methods are reasonably in accordance for the ligand: The SASA energy estimate is 9–16 kJ/mol (corresponding to SASAs of 240–550 Å^2), whereas the PCM non-polar energies are –3 to +24 kJ/mol with a correlation coefficient $r^2 = 0.85$.

This illustrates a major problem in estimating binding affinities using approaches that involve a continuum estimate of the solvation energy. Apparently, there is no consensus how the non-polar energy should be estimated and the PCM and SASA approaches give strongly differing results. It has previously been argued that it does not matter whether the area or volume is used to estimate the non-polar solvation energy [68]. However, the present results show that this is not the case for ligand-binding affinities: When a ligand binds to a complex, the volume of the protein increases, approximately by the volume of the ligand (so that the total volume during the binding reaction hardly changes). However, the SASA typically decreases during the binding, because the ligand becomes partly hidden by the protein and an empty cavity in the protein becomes filled by the ligand. In PCM, this is further complicated by use of several energy terms with different functional forms. In fact, it appears as the cavity term (after cancellation of the volume contributions) contains the same type of information as the SASA estimate (the difference is always within 3 kJ/mol), but that the dispersion and repulsion are either overestimated in PCM (because they rely on cancellation with the cavity term occuring for small molecules) or simply missing in SASA. At present, we cannot say which approach is more physical.

Another difference between the two solvation methods is that the PB method is based on SASAs, whereas the electrostatic and cavitation terms in PCM are based on the van der Waals surface of the solute. The latter is simply the surface of the union of spheres on all atoms with the corresponding van der Waals radius, whereas the SASA is defined as the surface defined by the centre of a spherical solvent probe that is rolled on the van der Waals surface. Therefore, the radius of a solvent molecule (~1.4 Å for water) is added to the van der Waals radii of each atom and thus, crevices between the spheres that are not accessible to a solvent probe are considered as a part of the solute. For small molecules, for which the PCM method was calibrated [50], these two surfaces are rather similar. However, for a large molecule, like a protein, they are totally different, because there are numerous small cavities inside the protein that are not large enough to room a solvent probe. The solvent-accessible surface of the protein will essentially be only the outer surface of the protein, whereas the van der Waals surface will be much larger. For example, for the avidin tetramer, the van der Waals area is 58 000 Å^2 and all atoms contribute to it, whereas the SASA is only 21 000 Å^2 and only 40 % of the atoms contribute to it.

It seems quite questionable to use the van der Waals surface to calculate any solvation energy in a protein, in particular for the electrostatic part of the solvation energy – surface charges are then calculated on many small boundaries inside the protein that should hardly be there. For the non-polar energies, we tend to prefer the SASA model, because it uses the seemingly better solvent-accessible surface, and additionally the resulting energies are 50 times smaller in magnitude, probably giving more precise differences. We have therefore based the recalibrated PCM model on the non-polar SASA energies. We do not argue that this is an optimum approach – on the contrary, it would be better to develop a new PCM method that works properly also for a protein, based on the solvent-accessible surface. Unfortunately, this is a major task, involving both method development and a complete reparametrisation of the method so that it works well both for small molecules and for proteins. Moreover, it has to be settled whether the non-polar term should be based on the volume or the surface area. Clearly, this is out of the scope of the present investigation.

Binding affinity estimates

Table 3 shows the various terms in the full PMISP/PCM/$T\Delta S$ method (with the non-bonded solvation energies from the PCM method; column ΔG_1). It can be seen that the method gives poor absolute affinities, ranging from +24 to +103 kJ/mol, compared to the experimental data, −19 to –85 kJ/mol. Therefore, the absolute errors for all the individual estimates are poor with a mean absolute deviation from the experimental estimates (MAD) of 105 kJ/mol. If we allow for a systematic error in the method (i.e. if we translated all points with the mean signed error), we still get a mean absolute deviation (TR MAD) of 21 kJ/mol, with the largest error for BTN1. This result is very disappointing. It is worse than similar MM/PBSA calculations using various MM force fields for the same system, which gave MADs of 9–19 kJ/mol, and TR MADs of 5–19 kJ/mol [33]. In particular, the standard MM/PBSA calculations for exactly the same snapshots, using the Amber 2002 force field give a MAD and TR MAD of 13 and 11 kJ/mol (Table 4). In fact, it is even worse than assigning the same affinity to all seven biotin analogues, which gives a TR MAD of 20 kJ/mol. The correlation coefficient is also quite poor, $R^2 = 0.22$, compared to 0.65 for MM/PBSA, and 0.43–0.98 in our previous investigation [33]. The replacement of the PCM non-polar term by the SASA term (as discussed above) gave only a slightly better TR MAD, 19 kJ/mol although the binding affinities are shifted to a range closer to the experimental one, –2 to –110 kJ/mol (column ΔG_2 in Table 3).

The standard deviations of the total PMISP/PCM/$T\Delta S_{MM}$ energies are listed in Table 5, where it can be seen that they are 10–30 kJ/mol. Thus, the standard errors of the mean values are 3–9 kJ/mol, showing that the statistical precision cannot explain the poor results. The standard deviation is dominated by the electrostatics, induction, polar solvation, and non-classical terms, which typically give slightly larger standard deviations than the total energy, because some of the variation between these terms is cancelled. The standard deviation of the entropy term is also quite large, 9–21 kJ/mol, but it never limits the precision of the method. The standard deviation of the non-polar solvation energy is always less than 1 kJ/mol. The corresponding standard deviations for the MM/PBSA method are also listed in Table 5. The standard deviations of the electrostatics and entropy terms are similar to that for PMISP, but those of the solvation and the non-classical terms are somewhat smaller.

We will try to rationalise the failure of PMISP by analysing the various terms in the method in comparison to MM/PBSA. The entropy term is identical between the two methods and the non-polar solvation term is also identical in the ΔG_2 estimate, so these cannot explain the failure.

The solvation energies show differences of –62 to +91 kJ/mol (PMISP mostly more negative for the neutral ligands and always more positive for the charged ligands). However, the correlation is excellent for the neutral ligands, $R^2 = 0.97$ and rather good for the charged ones, $R^2 = 0.93$.

The electrostatic and induction energies of the PMISP and MM/PBSA methods are not comparable, because intramolecular induction is not treated in the same way [19]. Therefore, we can only compare the sum of these two terms. It turns out that this sum is always more negative with PMISP than with MM/PBSA, by 39–47 kJ/mol for the neutral ligands, and by 135–232 kJ/mol for the charged ligands. However, again the two terms are almost perfectly correlated, with R^2 of 0.98 and 0.97 for the charged and neutral ligands, respectively.

If the solvation energy is added to this sum, the difference is partially cancelled, but the PMISP/PCM results are still 29–141 kJ/mol more negative than the MM/PBSA results. Unfortunately, the good correlation is completely lost, especially for the charged ligands ($R^2 = 0.12$, versus 0.87 for the neutral ligands).

Finally, the non-classical (van der Waals) energies also differ by a sizeable, but rather constant amount, 28–74 kJ/mol, which is slightly larger for the charged ligands than for the neutral ones. The PMISP estimates are always more positive. There is a perfect correlation ($R^2 = 1.00$) between Amber and PMISP for the neutral ligands, but it is much worse for the charged ones ($R^2 = 0.21$). If we replace the non-classical PMISP term with the Amber van der Waals term, the results become worse, with a TR MAD of 30 kJ/mol (but R^2 increases to 0.55).

Conclusions

In this paper we present the first attempt to calculate ligand-binding affinities using high-level QM methods with large basis sets (MP2/cc-pVTZ, i.e. enough to get reasonably accurate dispersion energies), combined with estimates of solvation energies, entropy, as well as sampling effects. To this aim, we have used the recently developed PMISP method [19], extended to a full protein, using the PMISP/MM approach [20]. This method uses an accurate MM force field, with multipoles up to quadrupoles and anisotropic polarisabilities in all atoms and bond centres, calculated for all residues in the correct conformation in the protein. This force field is then combined by explicit pair-wise QM calculations at the MP2/cc-pVTZ level for the ligand and all residues within 4 Å. This approach has been shown to give an accuracy of 5–15 kJ/mol compared to full QM calculations with the same method [19,20].

This should give accurate non-bonded interaction energies between the protein and the ligand. To obtain estimates also of other important terms in the ligand binding, we have combined the PMISP approach with the standard MM/PBSA method [4]. Thus, we supplement the PMISP energies with a continuum estimate of the solvation energy and an MM estimate of the entropy change during ligand binding, obtained from the vibrational frequencies. Moreover, all terms are calculated for ten snapshots from a MD simulation of the protein. Since the standard Poisson–Boltzmann and generalised Born methods cannot handle multipoles and polarisabilities, we have employed the PCM model, developed for the effective fragment potential [47].

Unfortunately, the results with this PMISP/PCM/$T\Delta S$ approach are poor in both absolute and relative terms, with a TR MAD of 21 kJ/mol, i.e. worse than a standard MM/PBSA method for the same problem (11 kJ/mol) and slightly worse than a trivial model predicting all binding affinities to be equal (20 kJ/mol). This can partly be caused by the use of different force fields for the geometry generation and for the energy calculations [33], but a more likely reason is that some error cancellation in the use of MM together with PB is lost when the interaction energies are improved and the solvent model is changed. We are currently investigating this issue further.

The reason for the poor absolute energies is probably the non-polar solvation energies, obtained with the PCM approach, which are 60–180 kJ/mol more positive than those obtained with the simple SASA-based method in standard MM/PBSA [69]. There are three important sources to this large difference. First, the cavity term in PCM is based on the van der Waals surface, which is ~3 times larger than the SASA for a protein. Second, the non-polar terms in PCM are ~50 times larger in magnitude than the corresponding SASA terms. Third, the cavity energy in PCM contains terms that are proportional to the change in the volume during ligand binding. Typically, the volume is essentially unchanged, whereas the surface decreases during ligand binding. Therefore, the net non-polar PCM term is positive, whereas the corresponding SASA term is negative (and ~5 times smaller). Moreover, it seems questionable to use a surface (e.g. the van der Waals surface) that contains contributions for atoms deeply buried in the protein and gives rise to many cavities. This might also explain the different results also in the polar solvation energies.

9

References

1   Gohlke, H.; Klebe, G.;*Angew. Chem. Int. Ed.* **2002**, *41*, 2644-2676.

2   D. L. Beveridge, F. M. Dicapua;1989, 18, 431-492.

3   T. Hansson, J. Marelius, J. Åqvist; *J. Comput.-Aided Mole. Design* **1998**, *12*, 27-35.

4   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, Acc. Chem. Res. 2000, 33, 889-897.

5   Raha, K.; Peters, M. B.; Wang, B.;Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M. Drug Discov. Today 2007, 12, 725-731.

6   Raha, K.; Merz, K. M.; J. Am. Chem. Soc. 2004, 126, 1020-1021.

7   Raha, K.; Merz, K. M.; J. Med. Chem. 2005, 48, 4558-4575.

8   Fukuzawa, K.; Mochizuki, Y.; Tanaka, S.; Kitaura, K.; Nakano, T.; J. Phys. Chem. B 2006, 110, 16102.

9   Nakanishi, I.; Fedorov, D. G.; Kitaura, K. Proteins, Struct. Funct. Bioinf. 2007, 68, 145-158.

10   Zhang D. W.; Xiang, Y.; Zhang, J. Z. H.; J. Phys. Chem. B 2003, 107, 12039-12041.

11   Zhang D. W.; Xiang, Y.; Gao, A. M.; Zhang, J. Z. H.; J. Chem. Phys. 2004, 120, 1145-1148.

12   Zhang D. W.; Zhang, J. Z. H.; Intern. J. Quant. Chem. 2005, 103, 246-257.

13   Mei, Y.; Xiang, Y.; Zhang D. W.; Zhang, J. Z. H. Proteins, Struct. Funct. Bioinf. 2005, 59, 489-495.

14   He, X.; Mei, Y.; Xiang, Y.; Zhang D. W.; Zhang, J. Z. H. Proteins, Struct. Funct. Bioinf. 2005, 61, 423-432.

15   Wu, E. L.; Mei, Y.; Han, K.; Zhang, J. Z. H. Biophys. J. 2007, 92, 4244-4253.

16   Bettens, R. P. A.; Lee, A. M. Chem. Phys. Lett. 2007, 449, 341-346.

17   Jurecka, T.; Sponer, J.; Cerny, J.; Hobza, P. Phys. Chem. Chem. Phys. 2006, 8, 1985.

18   Giese, T. J.; York, D. M. *J. Chem. Phys*. **2004**, *120*, 9903-9906.

19   Söderhjelm, P.; Ryde, U. J. Phys. Chem. B, 2009, accepted.

20   Söderhjelm, P.; Aquilante, F.; Ryde, U. J. Phys. Chem. A, submitted.

21   P. C. Weber, D. H. Ohlendorf, J. J. Wendolowski, F. R. Salemme, Science 1989, 243, 85-88.

22   P. C. Weber, J. J. Wendoloski, M. W. Pantoliano, F. R. Salemme, J. Am. Chem. Soc. 1992, 114, 3197-3200.

23   L. Pugliese, A. Coda, M. Malcovati, M. Bolognesi, J. Mol. Biol. 1993, 231, 698-710.

24   O. Livnah, E. A. Bayer, M. Wilchek, J. L. Sussman, Proc. Natl. Acad. Sci. U.S.A. 1993, 90, 5076-5080.

25   Green, N. M., *Biochem. J.* **1966,** *101*, 774-780.

26   N. M. Green, Avidin, Adv. Protein Chem. 1975, 29, 85-133.

27   N. M. Green, Avidin and streptavidin, Methods Enzymol. 1990, 184, 51-67.

28   Miyamoto, S.; Kollman, P. A. , Proteins, Struct. Funct. Genet. **1993**, *16*, 226-245.

29   Wang, J.; Dixon, R.; Kollman, P. A., Proteins, Struct. Funct. Genet. **1999**, *34*, 69-81.

30   B. Kuhn, P. A. Kollman, J. Med. Chem. 2000, 43, 3786-3791.

31   B. Kuhn, P. Gerber, T. Schulz-Gasch, M. Stahl, J. Med. Chem. 2005, 48, 4040-4048.

32   Brown, S. P; Muchmore, S. W., J. Chem. Inf. Model. 2006, 46, 999-1005.

33   Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U.; J. Med. Chem. 2006, 49, 6596-6606.

34   Gagliardi, L.; Lindh, R.; Karlström, G.; J. Chem. Phys. 2004, 121, 4494.

35   Zhang, D. W.; Zhang, J. Z. H. J. Chem. Phys. 2003, 119, 3599.

36   Riley, K. E.; Hobza, P. J. Phys. Chem. A 2007, 111, 8257.

37   Cornell,W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K. M.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. J. Am. Chem. Soc. 1995, 117, 5179-5197 .

38   Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee.; J. Comput. Chem. 2003, 24, 1999-2012.

39   P. Cieplak, J. Caldwell, P. Kollman.; J. Comput. Chem. 2001, 22, 1048-1057.

40   G. Karlström, R Lindh, P.-Å. Malmqvist, B. O. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrady & L. Seijo (2003) , Comput. Materials Sci., 28, 222-239.

41   Beebe, N. H. F.; Linderberg, J.; Int. J. Quant. Chem. 197, 7, 683.

42   Koch, H.; Sánchez de Merás, A.; Pedersen, T. B.; J. Chem. Phys., 2003, 118, 9481.

43   Aquilante, F.; Pedersen, T. B.; Lindh, R.; J. Chem. Phys. 2007, 126, 194106.

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

45   Schnieders, M. J.; Baker N. A.; Ren P.; Ponder, J. W. J. Chem. Phys. 2007, 126, 124114.

46   Schnieders, M. J.; Ponder, J. W. J. Chem. Theory Comput. 2007, 3, 2083-2097.

47   Li, H.; Pomelli, C. S.; Jensen, J. H. Thor. Chem. Acc. 2003, 109, 71-84.

48   Cances, E.; Mennucci, B.; Tomasi, J. J. Chem. Phys. 1997, 107, 3032

49   M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery; J. Comput. Chem., 1993, 14,

1347-1363.

50   Barone V.; Cossi, M.; Tomasi, J. J. Chem. Phys. 1997, 107, 3210-3221

51   Gaussian 03, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; 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.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004.

52   Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. J. Chem. Theory Comput. 2007, 3, 2011-2033.

53   Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. J. Phys. Chem. B 2006, 110, 16066-16081.

54   Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. Biochemistry 1981, 20, 849-855.

55   Sitkoff, D.; Sharp, K. A.; Honig, B. J. Phys. Chem. 1994, 98, 1978-1988.

56   Florián, J. Warshel, A. J. Phys. Chem. B 1997, 101, 5583.

57   Kongsted, J.; Söderhjelm, P.; Ryde, U.; J. Chem. Theory Comput., submitted.

58   Gilson, M. K; Honig, B., *Proteins, Struct. Funct. Gen.* **1998**, *4*, 7-18.

59   Hermann, R. B. , J. Phys. Chem. 1972, 76, 2754-2759.

60   Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A., *Biophys. J.* **2004**, *86*, 67-74.

61   Gräter, F.; Schwarzl, S. M.; Dejaegere, A.; Fischer, S.; Smith, J. C. J. Phys. Chem. B 2005, 109, 10474-10483.

62   Wang, M. L.; Wong, C. F. J. Chem. Phys. 2007, 126, 026101.

63   M. Kaukonen, P. Söderhjelm, J. Heimdal, U. Ryde, J. Phys. Chem. B, 2008, 112, 12537-12548.

64   Kongsted, J.; Ryde, U.; J. Comput. Aided Mol Design, published on-line; DOI: 10.1007/s10822-008-9238-z.

65   Sitkoff, D.; Sharp, K. A.; Honig, B. J. Phys. Chem. 1994, 98, 1978-1988.

66   Cossi, M.; Tomasi, J.; Cammi, R. Int. J. Quant. Chem. Quant. Chem. Symp. 1995, 29, 695-702.

67   Floris, F.; Tomasi, J.; J. Comput. Chem. 1989, 10, 616-627.

68   Tan, C.; Tan, Y.-H.; Luo, R.; J. Phys. Chem. B 2007, 111, 12263-12274.

69   Hermann, R. B. J. Phys. Chem. 1972, 76, 2754-2759.

Table 1. Calibration of the PCM method for PMISP. The total solvation free energy of 22 organic molecules and ions were calculated with the PCM+SASA method, using different values for the scaling factor of the radii for the electrostatic term (1.10–1.20). The SASA non-polar energy, calculated with Parse radii [11] were added to these values, and the results were compared to experiments. In the table, the difference compared to experiments are given, as well as the non-polar energy term (ΔE_{np}), and the experimental data (Exp.) [52,53,54,55,56] (all in kJ/mol).

	\multicolumn{11}{c}{Scaling factor of radii for polar term}	ΔE_{np}	Exp.
	1.20	1.19	1.18	1.17	1.16	1.15	1.14	1.13	1.12	1.11	1.10		
H_2O	7.9	7.1	6.1	5.2	4.1	3.1	1.9	0.7	-0.5	-1.9	-3.3	6.3	-26.4[a]
CH_3OH	9.8	9.1	8.3	7.5	6.6	5.7	4.7	3.6	2.5	1.3	0.0	7.4	-21.4[a]
ethanol	10.7	9.9	9.1	8.3	7.3	6.4	5.4	4.2	3.0	1.7	0.4	8.2	-21.0[a]
$p\text{-}CH_3C_6H_4OH$	9.9	8.8	7.7	6.5	5.2	3.9	2.4	0.9	-0.8	-2.6	-4.6	10.0	-25.7[a]
NH_3	8.4	7.9	7.3	6.7	6.1	5.4	4.7	4.0	3.2	2.3	1.5	6.5	-17.9[a]
CH_3NH_2	11.7	11.1	10.5	9.8	9.1	8.3	7.5	6.6	5.7	4.7	3.6	7.5	-19.1[a]
CH_3CONH_2	13.7	12.6	11.5	10.3	9.0	7.6	6.2	4.7	3.2	1.6	-0.2	8.3	-40.6[a]
propionamide	15.2	14.2	13.1	11.9	10.7	9.4	8.0	6.5	5.0	3.4	1.7	8.9	-39.2[a]
CH_4	-1.4	-1.4	-1.4	-1.4	-1.4	-1.4	-1.5	-1.5	-1.5	-1.6	-1.6	7.1	8.4[a]
propane	0.1	0.1	0.1	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.2	-0.2	8.5	8.2[a]
n-butane	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.1	0.1	0.0	9.3	8.7[a]
isobutane	-0.9	-0.9	-1.0	-1.0	-1.0	-1.1	-1.1	-1.1	-1.2	-1.2	-1.3	9.1	9.7[a]
toluene	5.9	5.5	5.2	4.8	4.4	4.0	3.5	3.0	2.5	1.9	1.3	9.6	-3.7[a]
CH_3SH	5.1	4.8	4.5	4.1	3.8	3.4	3.0	2.6	2.1	1.6	1.1	7.9	-5.2[a]
$CH_3SC_2H_5$	7.3	7.0	6.7	6.3	6.0	5.6	5.2	4.7	4.2	3.7	3.2	9.3	-6.2[b]
3-methylindol	8.2	7.1	6.0	4.8	3.6	2.3	0.9	-0.6	-2.2	-3.9	-5.7	10.7	-24.6[b]
4-methylimidazole	12.4	10.9	9.4	7.8	6.2	4.4	2.5	0.5	-1.6	-3.8	-6.3	9.1	-43.0[b]
N-propyl guanidine	17.3	15.5	13.7	11.7	9.6	7.4	5.0	2.5	-0.3	-3.3	-6.4	10.2	-45.7[c]
$CH_3NH_3^+$	36.8	34.2	31.6	28.9	26.1	23.3	20.4	17.4	14.4	11.3	8.1	7.8	-319.7[d]
imidazoleH$^+$	-9.2	-12.0	-14.8	-17.8	-20.9	-24.0	-27.3	-30.8	-34.3	-38.1	-42.0	8.5	-248.9[e]
$HCOO^-$	15.3	12.6	9.9	7.2	4.3	1.4	-1.6	-4.7	-7.9	-11.2	-14.6	7.2	-318.8[d]
CH_3COO^-	21.6	19.0	16.3	13.5	10.7	7.8	4.9	2.0	-1.0	-4.1	-7.3	8.1	-324.7[d]
MAD, all	10.4	9.6	8.8	8.0	7.1	6.2	5.4	4.7	4.4	4.8	5.2		
MAD, neutral	8.1	7.5	6.8	6.0	5.2	4.4	3.5	2.7	2.2	2.3	2.3		

[a] Data from [52].
[b] Data from [54].
[c] Data from [55].
[d] Data from [53].
[e] 18.8 kJ/mol was added to the value in ref. 56 to use the same value of the absolute solvation energy of a proton as in ref. no. 53.

Table 2. Average non-polar solvation energies (kJ/mol) in the SASA and PCM calculations. The energy contributions for the complex (*PL*), protein (*P*), and ligand (*L*) given, as well as the net contribution to the binding (*PL–P–L*), for PCM further divided into cavitation (cav), dipsersion (disp), and repulsion (rep) contributions.

	SASA				PCM						
	PL	*P*	*L*	*PL–P–L*	*PL*	*P*	*L*	*PL–P–L*	cav	disp	rep
BTN1	465	467	14	-16	22282	22156	12	114	-18	189	-57
BTN2	466	469	14	-17	22275	22144	14	116	-20	198	-61
BTN3	471	474	14	-17	22234	22117	7	111	-22	181	-49
BTN4	469	473	16	-20	22447	22268	24	155	-22	236	-58
BTN5	465	468	13 -	-16	22250	22138	8	104	-18	164	-41
BTN6	467	470	13 -	-16	22228	22119	9	100	-19	153	-34
BTN7	478	480	9 -	-11	21942	21892	-3	52	-8	89	-29

Table 3. The results of the PMISP/PCM/$T\Delta S$ method. Three different estimates of the total binding energy are given: $\Delta G_1 = \Delta E_{es} + \Delta E_{ind} + \Delta E_{nc} + \Delta G_{solvPCM} + \Delta G_{np,PCM} - T\Delta S$ is the full PMISP/PCM/TΔS, whereas in $\Delta G_2 = \Delta E_{es} + \Delta E_{ind} + \Delta E_{nc} + \Delta G_{solvPCM} + \Delta G_{np,SASA} - T\Delta S$, the non-polar PCM term has been replaced by the non-polar SASA term, and in $\Delta G_3 = \Delta E_{es} + \Delta E_{ind} + \Delta E_{vdW} + \Delta G_{solvPCM} + \Delta G_{np,SASA} - T\Delta S$, the ΔE_{nc} term has also been replaced by the Amber van der Waals energy. The mean absolute deviation (MAD), the correlation coefficient (R^2), as well as the MAD after subtraction of the mean signed deviation (MAD TR) are also given for each energy estimate.

Exp	ΔE_{es}	ΔE_{ind}	ΔE_{nc}	$\Delta G_{solv,PCM}$	$\Delta G_{np,PCM}$	$-T\Delta S$	$\Delta E_{np,SASA}$	ΔE_{vdW}	ΔG_1	ΔG_2	ΔG_3
BTN1	-1056.3	-253.5	-80.5	1236.5	113.9	96.8	-16.9	-143.4	57.0	-73.8	-136.7
BTN2	-1108.8	-322.7	-74.7	1311.3	116.3	102.4	-17.2	-149.1	23.9	-109.7	-184.1
BTN3	-1051.7	-282.3	-68.9	1235.0	110.7	93.4	*-16.7*	-138.6	36.2	-91.2	-160.9
BTN4	-124.4	-55.9	-150.2	181.1	155.4	96.5	-21.3	-211.0	102.6	-74.1	-134.8
BTN5	-113.2	-50.9	-95.8	140.7	104.5	77.4	-16.2	-134.5	62.6	-58.0	-96.7
BTN6	-91.5	-45.3	-91.4	132.9	100.4	69.8	-15.8	-131.9	74.9	-41.3	-81.8
BTN7	-113.3	-46.3	-25.1	127.4	52.3	66.4	-10.7	-53.6	61.2	-1.7	-30.1
MAD									104.7	27.5	72.9
TR MAD									20.6	19.2	30.2
R^2									0.22	0.50	0.55

Table 4. Results for the MM/PBSA calculations using the polarisable Amber 2002 force field. The MAD and TR MAD are 13.2 and 11.5 kJ/mol, respectively, and R^2 is 0.65.

	ΔE_{es}	ΔE_{ind}	ΔE_{vdW}	$\Delta E_{solv,PB}$	$\Delta E_{np,SASA}$	$-T\Delta S$	ΔG_{bind}	Exp.
BTN1	-1173.6	-1.6	-143.4	1180.0	-16.9	96.8	-58.6	-85.4
BTN2	-1213.8	14.6	-149.2	1220.0	-17.2	102.4	-43.2	-59.8
BTN3	-1182.2	-0.6	-138.6	1164.1	-16.7	93.4	-80.6	-58.6
BTN4	-127.0	-14.6	-211.0	243.1	-21.3	96.6	-34.2	-36.8
BTN5	-100.7	-16.3	-134.5	159.4	-16.1	77.4	-30.9	-34.3
BTN6	-80.4	-11.2	-131.9	152.1	-15.8	69.8	-17.4	-20.9
BTN7	-111.4	-9.6	-53.6	117.6	-10.6	66.4	-1.2	-18.8

Table 5. Standard deviations of the various terms for PMISP/PCM/$T\Delta S_{MM}$ and MM/PBSA. E_{eis} is the sum of the E_{es}, E_{ind}, and G_{solv} terms.

	MM/PBSA								PMISP/PCM/$T\Delta S_{MM}$						
	E_{es}	E_{ind}	E_{vdW}	G_{PB}	G_{SASA}	$T\Delta S$	G_{bind}	E_{eis}	E_{es}	E_{ind}	E_{nc}	G_{PCM}	$G_{np,PCM}$	G_{bind}	E_{eis}
BTN1	20.9	5.4	15.0	18.8	0.2	14.2	18.7	19.6	18.1	17.6	24.8	22.0	0.2	23.3	18.5
BTN2	37.6	4.6	14.4	21.0	0.2	21.0	25.4	25.5	33.3	32.2	22.7	45.4	0.2	29.8	35.5
BTN3	26.0	6.8	11.6	17.9	0.2	11.7	22.1	29.3	25.9	24.2	25.3	24.8	0.2	12.5	30.8
BTN4	15.5	3.8	10.8	16.5	0.3	11.8	20.7	19.8	19.7	9.8	14.0	11.9	0.3	17.0	20.8
BTN5	18.9	3.9	8.0	16.7	0.1	12.1	21.8	23.6	13.5	13.3	13.6	15.8	0.1	10.5	15.0
BTN6	15.1	3.1	15.3	8.7	0.2	15.3	21.1	17.4	13.3	11.0	19.2	15.4	0.2	12.8	14.1
BTN7	10.6	3.1	7.8	10.9	0.1	9.3	18.5	16.6	11.4	6.9	14.5	8.4	0.1	11.8	11.9