LUND UNIVERSITY

How to obtain statistically converged MM/GBSA results.

Genheden, Samuel; Ryde, Ulf

Published in:
Journal of Computational Chemistry

DOI:
10.1002/jcc.21366

2010

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

Link to publication

Citation for published version (APA):
Genheden, S., & Ryde, U. (2010). How to obtain statistically converged MM/GBSA results. *Journal of Computational Chemistry*, *31*(Online 13 Jul 2009), 837-846. https://doi.org/10.1002/jcc.21366

Total number of authors:
2

Creative Commons License:
Unspecified

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

How to obtain statistically converged MM/GBSA results

Samuel Genheden, 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

2017-03-12

Abstract

The molecular mechanics/generalized Born surface area (MM/GBSA) method has been investigated with the aim of achieving a statistical precision for the results of 1 kJ/mol. We studied the binding of seven biotin analogues to the avidin, taking advantage of the fact that the protein is a tetramer with four independent binding sites, which should give the same estimated binding affinities. We show that it is not enough to use a single long simulation (10 ns), because the standard error of such a calculation underestimates the difference between the four binding sites. Instead, it is better to run several independent simulations and average the results. With such an approach, we obtain the same results for the four binding sites, and any desired precision can be obtained by running a proper number of simulations. We discuss how the simulations should be performed to optimise the use of computer time. The correlation time between the MM/GBSA energies is ~5 ps and an equilibration time of 100 ps is needed. For MM/GBSA, we recommend a sampling time of 20–200 ps for each separate simulation, depending on the protein. With 200 ps production time, 10–80 separate simulations are required to reach a statistical precision of 1 kJ/mol (1600–12800 energy calculations or 3–24 ns total simulation time per ligand) for the seven avidin ligands. This is an order of magnitude more than what is normally used, but such a number of simulations is needed to obtain statistically valid results for MM/GBSA method.

Key Words: MM/PBSA, ligand-binding affinities, generalised Born, avidin, biotin, galectin

Introduction

In biochemistry, most small molecules exert their action by binding to a macromolecule. Therefore, it is of great importance to understand the reaction

$$L + R \rightleftharpoons LR \quad (1),$$

where L is the ligand, R is the macromolecular target (the receptor) and LR is the complex. In particular, a main goal of computational medicinal chemistry is to develop methods that accurately can estimate the free energy of this reaction, ΔG_{bind}, as this would allow us predict the binding strength of any drug candidate without synthesising it. Indeed, this has been described as the Holy Grail of structure-based drug design [1].

The most accurate and rigorous methods for ligand binding are free energy perturbation (FEP) [2] and thermodynamic integration (TI) [3]. They are founded on statistical mechanics, and are well-established and in theory exact [4], In these methods, the difference in binding free energy between two ligands is calculated by slowly changing one ligand into another, via a number of unphysical, intermediate states, using molecular dynamics (MD) or Monte Carlo simulations. Despite initial success and promising results, FEP or TI has found relatively little use in drug design [5], because they converge only for rather similar ligands and are very computational expensive. As such, they are usually only applicable to congeneric series of inhibitors [1,4].

Therefore, more simplified and faster methods have been developed, such as the linear interaction energy [6] and the molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) [7] methods. These two methods only simulate the end-points of the reaction, i.e. only physical states. There are also other methods that do not require any simulation at all and relay on statistical relationships [1], but such methods do not give accurate and uniform results for all types of complexes [1,8].

In this work we have concentrated on MM/PBSA and the related MM/GBSA method [7]. They are promising methods that have been widely used and have given good results [9,10,11], although there are also cases reported for which they have worked less satisfactorily [12,13,14]. In both approaches, the binding free energy is approximated by the difference in free energy of the three reactants:

$$\Delta G_{bind} = \langle G_{RL} \rangle - \langle G_R \rangle - \langle G_L \rangle \quad (2),$$

where each free energy is estimated as a sum, according to

$$G = E_{MM} + G_{solv} - TS_{MM} \quad (3),$$

where G_{MM} is the molecular mechanics gas-phase energy of the reactant, consisting of the internal energy (from bonds, angles, and dihedral angles), as well as the non-bonded electrostatic and van der Waals energies:

$$E_{MM} = E_{int.} + E_{el.} + E_{vdW} \quad (4)$$

G_{solv} is the solvation energy, and is calculated with a continuum representation of the solvent for the polar part, and by a relation to the solvent accessible surface area for the non-polar part. The polar part can either be calculated by solving the Poisson–Boltzmann equation [15] (giving MM/PBSA [7]) or by using the generalized Born method [16] (MM/GBSA). The last term TS_{MM} is the product of the absolute temperature and the entropy, which is calculated from a normal-mode analysis of a truncated system at the molecular-mechanics level.

The averages in Eqn. 1 are calculated from a set of snapshots taken from a molecular dynamics simulation to include the effects of dynamics. Each of the three free energies in

Eqn. (2) should in principle be calculated from an individual simulation. However, it is more common to only simulate the complex and then calculate all three free energies in Eqn. 2 from this simulation [17,18]. In that case, the internal MM energy (E_{int}) cancels out.

A major problem with the MM/PBSA approach is the large standard deviation of the estimate of the binding free energy, coming from a large variation in the individual terms in Eqns. 3–4 among the snapshots from the MD simulation. Unfortunately, this estimate is often not reported, but it is typically 20–150 kJ/mol (and 120–210 kJ/mol if separate trajectories are used for the three reactants) [13,19,20,21]. Considering that drug candidates often differ in their affinity by less than order of magnitude in the binding constant, corresponding to 6 kJ/mol, this is a major problem. Of course, it can be cured by performing more energy calculations, but if you aim at an accuracy of half of that (close to the experimental accuracy) and want to draw statistically significant conclusions (the *t* distribution value is ~2 at 95 % confidence), the desired standard error should to be close to 1 kJ/mol. Hence, 400–22500 energy calculations are needed. This is much more than what is traditionally used (10–200), but a few recent studies have used up to 10 ns simulation time and 1000 energy calculations (but only up to 50 for the demanding entropy calculations) [13,19,20]. However, these investigations have also pinpointed another problem, namely the occurrence of several substates that are only seldom sampled during the simulations, but may give binding energies that differ by an amount larger than what is expected from the standard error [20], indicating that even 10 ns are too short time to obtain truly equilibrated results.

In this paper we thoroughly investigate these issues, examining what is needed to obtain MM/GBSA results that have a true statistical accuracy of ~1 kJ/mol. To examine whether the results are statistically valid, we have selected a test case for which there are four equivalent binding sites in the same protein, viz. the binding of seven biotin analogues (Figure 1) to avidin. This protein is well-characterized by X-ray crystallography [22,23,24,25] and experimental binding affinities are available [26,27,28]. This system has been the subject of several studies with FEP [29,30], LIE [30] and MM/PB(GB)SA [12,21,31,32]. For this protein, we can calculate four independent binding constants from each simulation. If the results are converged, all four estimates should be identical, within the statistical uncertainty.

Methods

Preparation of complexes

We have studied the binding of the seven biotin analogues in Figure 1 to avidin. The preparation of the avidin protein and the inhibitors has been describe before [21]. The Amber 99-SB force field [33] was used for the protein atoms and the inhibitors were described with the Amber 99 force field [34] with charges [21] derived from RESP (restrained electronic potential) calculations [35]. Each protein–ligand system was immersed in an octahedral box of TIP4P-Ewald [36] water molecules that extended at least 10 Å outside the protein. The Amber 99-SB force field, especially in combination with TI4P-Ewald water, has in several investigations been shown to give improved structures and dynamic properties compared to experiments [33,37,38,39].

The preparation of galectin-3 with a substituted benzamido-lactosamine inhibitor (ligand **3** in [40], shown in Figure S1 in the supplementary material) has not been described before. The calculations were based upon a crystal structure of galectin-3 in complex with this ligand (PDB code 1kjr) [40]. The ligand was described with the generalized Amber force field 41 and charges were derived from RESP calculations [34], based on electrostatic potentials calculated at the HF/6-31G* level according to the Merz–Kollman scheme [42], but using a higher than default density of points (10 concentric layers with 17 points/Å^2). The protein was described with the Amber 99-SB force field [33]. Protons were added to galectin-3 assuming standard protonation states at pH 7 for all residues. The protonation state of the histidine

residues was decided from the local surroundings and hydrogen-bond networks: residue 158 was protonated on $N^{\delta 1}$ and the residues 208, 217 and 223 were protonated on $N^{\varepsilon 2}$. This gives a total charge of +4 on the system. The protein–ligand complex was solvated in an octahedral box of TIP4P-Ewald water molecules extending at least 8 Å from the protein.

Simulation protocol

All simulations were run by the sander module in Amber 10 [43]. The SHAKE algorithm [44] was used to constrain bond lengths involving hydrogen atoms. The temperature was kept at 300 K using Langevin dynamics [45] with a collision frequency of 2.0 ps^{-1}. Particle-mesh Ewald (PME) [46] with a fourth-order B-spline interpolation and a tolerance of 10^{-5} was used to treat long-range electrostatics. The non-bonded cut-off was 8 Å and the non-bonded pair list was updated every 50 fs. The MD time step was 2 fs.

The complex was first optimised by 500 steps of steepest descent minimisation, keeping all atoms, except water molecules and hydrogen atoms, restrained to their start position with a force constant of 418 kJ mol^{-1} $Å^{-2}$. The minimisation was followed by 20 ps MD equilibration with a constant pressure and the restraining force constant reduced to 214 kJ mol^{-1} $Å^{-2}$. Finally, a production simulation was run without any restraints, but still with a constant pressure. The length of this simulation was varied, as was the length of the initial discarded equilibration period.

MM/GBSA calculations

ΔG was calculated according to Eqns. 2–4. All terms in Eqn. 4 was calculated with Amber 8 [47] with all water molecules stripped off and with an infinite cut-off. The GB^{OBC} method (with [?], [?], and [?] set to 0.8, 0, and 2.91, respectively) and with the default modified Bondi radii was used to calculate the polar solvation energy [48], and the non-polar solvation energy was calculated using the solvent accessible surface area, according to

$$[?]G_{np} = [?]\,\mathrm{SASA} + b \tag{5}$$

with [?] = 0.0227 kJ/mol/$Å^2$ and b = 3.85 kJ/mol [31]. The entropy was calculated by a normal-mode analysis of the harmonic frequencies calculated at the MM level. For this calculation, we used our recently described modification to increase the precision [49]: All residues more than 12 Å from any atom in the ligand were deleted and the remaining atoms were minimised, keeping all residues more than 8 Å from the ligand (including all water molecules) fixed, to keep the geometry as close as possible to the original structure. Thereby, the questionable use of a distance-dependent dielectric constant can be avoided. In the frequency calculations, the fixed buffer region was omitted. The energy was calculated for each subunit in avidin, and the inhibitors in the other subunits were treated as a part of the protein. Prior to the MM/GBSA calculations, the ligand in the subunit of interest was centred in the octahedral box.

Estimation of correlation time

To ensure that all data in our statistical analysis are uncorrelated, we estimated the correlation time of the MM/GBSA energy estimates with the statistical inefficiency method [50,51]. Thus, we calculated the measure:

$$\Phi = \frac{\tau \cdot \sigma^2(Y)_\tau}{\sigma^2(X)} \tag{6}$$

where $\sigma^2(X)$ is the variance of the distribution $\{X\}$, i.e. the variance of the MM/GBSA ?G_{bind} estimates for the various snapshots and $\sigma^2(Y)$ is the variance of the block average of $\{X\}$, where the block length is τ. This block average is calculated from

$$Y_i = \frac{1}{\tau} \cdot \sum_{j=n-i\tau+1}^{n-(i-1)\tau} X_j \qquad (7)$$

That is, $\{X\}$ is divided into a number of segments, each with length τ. Once τ is so large that the successive values of Y_i are statistical independent, Φ will become a constant and an estimate of the correlation time of $\{X\}$.

Error estimates

To quantify the performance of the MM/GBSA method, we use three different estimates: the correlation coefficient between the predicted and experimental data (r^2), the predictive index (PI) [8], and the mean absolute deviation (MAD) from the best correlation line through the origin (i.e. after the subtraction of the mean signed difference). These estimates are quite meaningless without an estimate of their statistical uncertainty. They were obtained by a simple simulation: Each inhibitor was assigned a random number from a normal distribution, with the mean and standard deviation obtained in the MM/GBSA calculations. We then calculated MAD, r^2 and PI, and repeated this procedure 10 000 times. The standard deviations within these three sets are reported as the standard error of the error estimates. Throughout this paper, all reported statistical uncertainties are standard errors of the mean, i.e. the standard deviation divided by the square root of the number of estimates.

Result and Discussion

One long simulation

First, we examined whether it is enough to run a single long simulation. We run a MD simulation of the avidin tetramer with four bound biotin ligands (Btn1 in Figure 1) for 10 ns (after 0.2 ns equilibration), and calculated MM/GBSA ?G_{bind} estimates every picosecond. Statistical inefficiency calculations indicated that the correlation time for ?G_{bind} was ~5 ps, giving 2000 energy estimates for each subunit. The results are collected in Table 1. The standard deviation of ?G_{bind} is ~20 kJ/mol, so the standard errors of the mean values (the standard deviation divided by ?2000) are 0.5 kJ/mol, indicating that the energies are well converged.

However, the binding energies of biotin to each of the four subunits show a rather large variation, –110 to –121 kJ/mol. This is much larger than expected from the standard errors – all differences larger than 2.3 kJ/mol are 99.9% statistically significant. Of course, this could indicate that the true binding strengths of the four ligands are different, but there is no experimental support of this. Instead, a more plausible explanation is that the estimated standard errors are too low or that the conformational space is not sampled enough to obtain converged results. Calculating backwards from the standard deviations of the individual simulations, the large difference in ?G_{bind} indicates that the number of independent observations must be less than ~35 (i.e. a correlation time of 280 ps). This is a too large difference to assign the error to the estimate of the correlation time. Instead, the estimate of the standard deviation must be too small, which is equivalent of saying that the conformational space is not sampled enough. This is a very serious observation, indicating that 10 ns simulation time is far to small to obtain a proper sampling of the phase space and to obtain converged estimates of binding energies with the MM/GBSA method. Again

calculating back from the observed difference, the results indicate that a ~30 times longer simulation is needed to obtain converged results (i.e. 300 ns), which would seem to be prohibitively long.

Several short simulations

If we assume that the problem with the long simulation is that it remains too close to the starting structure and does not properly sample the full phase space, a possibly more economical solution to the problem could be to run several shorter simulations starting at different points in the phase space. A simple way to obtain this is to use the same starting structure, but using different starting velocities. This can very easily be obtained by simply providing a different seed to the random number generator, because the starting velocities are assigned by random to a Maxwell distribution. A similar approach has been used frequently before to obtain independent simulations [52,53,54].

Therefore, we run 20 independent 400 ps long simulations of biotin in avidin. Again, the correlation time for the MM/GBSA energies was ~5 ps and only the last 200 ps of the simulation were used (giving 800 energy estimates per subunit). The results are summarised in Table 2. The standard errors are now calculated simply from the standard deviation in the 20 independent simulations. It can be seen that they are larger than for the single simulation, 1.5–2.9 kJ/mol. On the other hand, the ΔG_{bind} estimates for the four subunits are more similar,
–111 to –117 kJ/mol. In this case, the largest difference (between subunits B and C) is not larger than what can be expected with the observed standard errors (90% significant difference, obtained for one of six possible pairs of subunits). Therefore, the total average over the four subunits and the 20 simulations, –114 kJ/mol, should be a proper estimate of the binding energy for biotin to avidin, using this force field, with a standard error of 1.1 kJ/mol (based on 80 observations). Of course, if we increase the number of independent simulations, the error will decrease.

Figure 2 shows the individual MM/GBSA estimates for the four subunits in the 20 short simulations. It can be seen that there is a large spread of the estimates, up to 55 kJ/mol (–81 to –135 kJ/mol), even if they are based on an average over 40 snapshots each. This indicates that we must use a large number of estimates to obtain converged results. It is also clear from the plot, that there is not apparent difference between the subunits, which shows that we indeed can treat them as independent. Thus, we seem to have solved the problem with sampling of the phase space.

Designing an practical procedure

We can now turn to the problem of designing a practical procedure that gives converged MM/GBSA results, i.e. to discuss the sampling rate, the length of the short simulations, and the equilibration time needed. The goal is to obtain as good results as possible with a minimum use of computer time.

We start with considering the sampling frequency. Therefore, the correlation time of the MM/GBSA energy was estimated using the statistical inefficiency method for all the four subunits in the 20 short simulation of the biotin–avidin complex. The results are shown in Figure 3, both for the whole simulated time (400 ps) and for only the last 200 ps of the simulation. It can be seen that the correlation time is 1–16 ps if all data is included and 1–7 ps if 100 ps equilibration time is excluded. Thus, the data indicates that it is favourable to exclude the first part of the simulation. In both cases, the most common correlation time is 2 ps. Based on these data, we decided to use a sampling frequency of 5 ps, for which 98 % of the data are uncorrelated if the equilibration period is excluded. Of course, this selection is somewhat arbitrary; we could have selected 7 ps, for which all 80 samples are uncorrelated,

but this would have been a waste of computer time by a factor of two for 94% of the simulations and for other ligands it is likely that larger correlation times can be observed. A correlation time of 3 or 4 ps could also be considered, for which 94 and 96% of the samples are uncorrelated, but we decided to use the more conservative measure of 5 ps to be on the safe side also for other systems.

The next step is to establish how long equilibration time that needs to be excluded to get converged results. Figure 4 shows how the MM/GBSA ΔG_{bind} energy varies with the equilibration time. Several different lengths of the production simulation was tested, but it can be clearly seen that the energies in all cases are converged (within 0.6 and 0.2 kJ/mol for 50 and 200 ps production time, respectively) when the equilibration time is 100 ps or longer.

Finally, we should establish how long production simulation is needed. Figure 5 shows the MM/GBSA ΔG_{bind} energy as a function of the length of the production simulation, for a number of different equilibration times. Again, we see that an equilibration time of 100 ps is needed for stable results. With this equilibration time, it can be seen that the results are converged to within 0.2 kJ/mol already after 75 ps. However, the curve for 200 ps equilibration shows that this is somewhat fortuitous – in that simulation, convergence (within 0.3 kJ/mol) is not reached until 175 ps. Therefore, we suggest a production simulation time of 200 ps.

Of course, it is possible that the suggested times strongly depend on the simulated system (both ligand and receptor). To check this, we have repeated the calculations also for another ligand of avidin, Btn2 in Figure 1. The results of these calculations are shown in Figures S2–S4 in the supplementary material. It can be seen that there are some variations, e.g. a slightly longer correlation time for the unequilibrated data, a somewhat longer equilibration time, and a slower convergence with respect to the length of the production time. However, the suggested values (5 ps sampling frequency, 100 ps equilibration time, and 200 ps production time) are still appropriate.

Moreover, to test the transferability of this procedure also to other systems, we carried out similar calculations on galectin-3 in complex with a substituted benzamido-lactosamine inhibitor (Figure S1) [40]. The results of these calculations are show in Figures S5–S7. For this protein, we had initially large problem to obtain equilibrated results. In all simulations, we obtained large changes in the MM/GBSA energies even after 300 ps equilibration. However, if we first run one long simulation (10 ns), and then started all independent simulations with different starting velocities from the end of that simulation, we actually obtained equilibrated data already after ~40 ps. This is probably a proper approach for any protein to save simulation time, i.e. to first run one long equilibration, before starting the independent simulations (with shorter equilibration times). It is likely that it is enough to run the initial equilibration for only ~1 ns, which would lead to a net gain in computer time. Besides this, the same simulation protocol is appropriate also for this protein.

Efficiency

Figure 6 shows the standard error of the MM/GBSA ΔG_{bind} energy as a function of the length of the production simulation for the biotin–avidin system. It can be seen that for 100 ps equilibration time, the standard error has not fully stabilised even after 200–300 ps of production time. The reason for this is that the reported standard deviation is calculated only for the 80 independent results for the four subunits in the 20 short simulations. However, each ΔG_{bind} estimate from the short simulations is also an average over a number of snapshots, determined by the length of the production simulations (divided by the sampling frequency of 5 ps). Therefore, the effective standard error is reduced as the length of the production simulations are increased, but not fully by $\sqrt{n}$ (where n is the number of energy calculations), because the results are not fully uncorrelated. This is illustrated in Table 3, which shows how much the standard deviation is reduced when the number of energy calculations (i.e. the

length of the production simulation) is increased in the short simulations, compared to the expected dependence of $n^{-½}$. It can be seen that already after two energy calculations, only 71% of the expected decrease is obtained, and after 41 energy calculations (200 ps production simulation), only 16% of the expected efficiency is obtained.

To decide the optimum procedure, we need to consider the relative cost of the molecular dynamics simulation and the MM/GBSA energy calculations. With the present simulation protocol and 3.0 GHz Intel Xenon 5160 processors, a single energy calculation takes ~2 CPU hours, whereas a 100 ps MD simulation takes 16 h. Thus, for our tetrameric protein, for which we can calculate four independent energies for each snapshot, the first energy calculation (after 100 ps equilibration) costs 16 (MD) + 8 (energy) = 24 CPU h. Each successive set of four energy calculations costs 0.8 (5 ps MD) + 8 = 8.8 CPU h. Given these estimates, we can now for any given total CPU time (CPU_{total}) calculate the estimated standard error of the final ?G_{binf} estimates (s_{av}) based on the actual standard deviations in Table 3 (s_{simu}) and the CPU cost of each individual simulation ($CPU_{simu} = 24 + 8.8(n - 1)$), according to

$$s_{av} = \frac{s_{simu}}{\sqrt{CPU_{tot}/CPU_{simu}}} = \frac{s_{simu}\sqrt{CPU_{simu}}}{\sqrt{CPU_{tot}}} \quad (7)$$

Thus a minimum of $s_{simu}\sqrt{CPU_{simu}}$ will indicate the most effective method. This estimate is also shown in Table 3. It can be seen that the optimum use of computer time is obtained after six energy calculations (25 ps production time).

However, for the more typical case of a single ligand site per receptor (in which case, the first energy calculation costs 18 CPU h and the successive cost 2.8 CPU h), it is more favourable to have a production simulation time of 40 ps (9 energy calculations). For very large proteins, for which the MD simulation cost is much higher than that of the energy calculation, a larger number of energy calculations per simulation is more favourable. On the other hand, if the energy calculations are much more expensive, e.g. quantum chemical calculations of the whole or part of the protein, which recently have started to emerge [55,56,57], it is clearly more favourable to use only a single energy calculation per simulation.

This reasoning clearly shows that the optimum length of the production simulation depends on details of the simulated system and should be recalculated for each system. However, it seems clear that typically it should be quite short (20–40 ps), at least for efficiency reasons (but of course, then a larger number of independent simulations need to be run in order to get a low standard error).

Seven biotin analogues

To illustrate our approach, we have estimated the binding affinities of all seven biotin analogues in Figure 1 to avidin. We used the protocol designed above, with a sampling frequency of 5 ps, an equilibration time of 100 ps, and a production time of 200 ps (which even if it is not the most efficient choice, gives converged results, according to Figure 5). The results are collected in Table 4.

When using 20 independent simulation, three of the inhibitors had a standard error of less than 1 kJ/mol (in fact, only 5–14 separate simulation would have been needed to converge these calculations), whereas it was slightly more for the other four. Therefore, we run some additional simulations to decrease the standard error below 1 kJ/mol for all seven ligands. From Table 4, it can be seen that up to 50 simulations were needed.

Figure 7 shows the correlation between the calculated and the experimental binding energy. It can be seen that the calculated affinities are systematically too negative by ~39 kJ/mol. This most likely comes from the continuum solvation model, which is quite crude and often gives systematic errors of this size [19,58]. In fact, we have previously shown that for the biotin–avidin system, the Poisson–Boltzmann method gives more accurate absolute

estimates [21], although for other systems, the opposite is true [49]. We have here selected to use the GB model, because it is much cheaper than the Poisson–Boltzmann method and gives more stable [49] and often even more accurate solvation energies [58]. In addition, a $3RT$ term is missing in the standard MM/PB(GB)SA estimate [19], making the absolute values dubious. Therefore, the relative results are much more interesting than the absolute ones, and then most solvation methods are expected to give similar results (within ~4 kJ/mol) [58].

However, from Figure 7, we can see that also the relative values are not very accurate: The mean absolute deviation from the relative correlation line (i.e. after the subtraction of the mean signed difference; MAD) is 15 kJ/mol. Likewise, the correlation coefficient r^2 is 0.59 and the predictive index is 0.85. This is caused by limitations in the MM/GBSA method and in the force field used. However, this is no problem in this investigation, because our aim is only to provide results that are statistically converged to 1 kJ/mol, i.e. precise results – we do not attempt to reduced the systematic (non-statistical) errors and also obtain accurate results.

In fact, we claim that we for the first time have obtained a statistically valid estimate of the true accuracy of a MM/GBSA estimate. With a simple random simulation (see the Methods section), we can obtain error estimates of the predicted MAD, 15 ? 0.3 kJ/mol, as well as for the correlation coefficient and the predictive index: $r^2 = 0.59$? 0.01 and PI = 0.85 ? 0.03.

The advantage with the present approach is that we now can compare these results with results obtained with other methods or with other force fields in a statistically valid way, and therefore discuss which method gives the more accurate results. This has hardly been possible before. For example, in our previous MM/PBSA study of the same biotin–avidin system with different simulation methods and force fields [21], we had a standard error for the calculated binding affinities of 10–14 kJ/mol (averages over 20 snapshots). Therefore, the standard deviations of the estimated MADs was ~3 kJ/mol, indicating that the MADs must differ by more than ~6 kJ/mol to be statistically significant to 95 % (this was also realised in a qualitative way in the article). Likewise, the simulated standard deviation of r^2 estimates is ~0.12, indicating that r^2 must differ by 0.24 to indicate a significant difference. This made MM/PBSA a very blunt method to compare different force fields. The situation was somewhat improved by our introduction of a more precise estimate of the entropy, which reduced the standard deviation of the ? G_{bind} estimates from ~50 to ~25 kJ/mol [49], but the standard deviation of the MAD estimate is still 2 kJ/mol if only 20 snapshots are used. Moreover, the present results indicate that these error estimates are too low, because only a single simulation was used. With the present methods, we can obtain statistically converged results with a precision that allow statistically valid comparisons between various methods and force fields.

Conclusions

In this work we have investigated how to obtain converged results with the MM/GBSA method. First, we have shown, taking advantage of the fact that there are four independent binding sites in avidin, that a single long simulation does not give converged results even after 10 ns. Instead, several short and independent simulations are needed. This is a very important finding, indicating that MD results are correlated, even after 10 ns.

Second, we have devised a MM/GBSA procedure to obtain a statistical precision of 1 kJ/mol. We suggest that a proper number of independent MD simulations should run, using different starting velocities. The final ? G_{bind} estimates and standard errors should be calculated from these series alone and the number of simulations is chosen to obtain the desired final accuracy (standard error of the mean value of ? G_{bind}, which is inversely proportional to the square root of the number of simulations). An equilibration of ~100 ps needs to be run before any data is collected. Moreover, time can often be saved if several energy estimates are calculated for each short simulation (because of the equilibration time). The correlation time of the MM/GBSA estimates are typically 1~7 ps and we suggest a

sampling frequency of 5 ps, to ensure that the individual data are uncorrelated. The optimum distribution between the number of short simulations and the number of energy calculations in each short simulation depends on the details of the system. In our test calculations, we obtain converged results with 5–50 independent simulations and 40 energy calculations (for each subunit) in each. However, efficiency calculations indicate that it would be more favourable in terms of CPU time to perform only six energy calculations in each independent simulation and instead run ~4 times more independent simulations. For monomeric proteins, more energy calculations (~9) are probably better, whereas for very time-consuming energy methods (e.g. quantum chemical calculation) a single energy calculation per independent simulation is more economic.

Using this approach, we have obtain binding affinities for the seven biotin analogues in Figure 1 to avidin, with standard errors of less than 1 kJ/mol. This required 1.5–15 ns simulation time and 800–8000 MM/GBSA energy calculations, depending on the ligand. This is more than what is normally used in MM/PB(GB)SA, but it is what is needed to obtain converged results.

Using these results, we can also judge how well the calculations reproduce the experimental binding energies in a statistically valid way. For the present MM/GBSA simulation protocol and the Amber 1999-SB force field with TIP4P-Ewald water molecules, we obtain a relative MAD of 15 ? 0.3 kJ/mol, a correlation coefficient of 0.59 ? 0.01, and a predictive index of 0.85 ? 0.03. The GB^{OBC} continuum solvation model gives systematically too negative absolute binding energies by 39 ? 1 kJ/mol. Even if we have used the currently best Amber force field (at least to reproduce structural properties) [33,36], this is not very impressive results. However, the reproduction of experimental data is not the aim of the present investigation. Instead, we provide for the first time statistically fully converged MM/GBSA results and statistical estimates of how well the method reproduce experimental data. Only with such estimates will it be possible to compare various methods and force fields in a statistically valid way, both in terms of efficiency and accuracy.

Acknowledgements

This investigation has been supported by grants from the Swedish research council and from the Research school in pharmaceutical science. It has also been supported by computer resources of Lunarc at Lund University and HPC2N at Umeå University.

References

1 Gohlke, H., Klebe, G. (2002) Angew. Chem., Int. Ed. 41:2644-2676
2 Zwansig, R. W. (1954) J. Chem. Phys. 22:1420-1426
3 Kirkwood, J. G. (1935) J. Chem. Phys. 3:300-313
4 Simson, T., Archontis, G., Karplus, M. (2002) Acc. Chem. Res. 35:430-437
5 Chipot, C., Rozanska, X., Dixit, S. B. (2005) J. Comput.-Aided Mol. Des. 19:765-770
6 Åqvist, J., Medina, C., Samuelsson, J. E. (1994) Prot. Eng. 7:385-391
7 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. (2000) Acc. Chem. Res. 33:889-897
8 Pearlman, D. A., Charifson, P. S. (2001) J. Med. Chem. 44:3417-3423
9 Fratev, F., Jonsdottir, S. O, Mihaylova, E., Pajeva, I. (2009) Mol. Pharmaceutics. 6:144–157
10 Grazioso, G., Cavalli, A., de Amici, M., Recanatini, M., de Micheli, C. (2008) J. Comput. Chem. 29:2593-2603
11 Fogolari, F., Moroni, E., Wojciechowski, M., Baginski, M., Ragona, L., Molinari, H.

(2005) Proteins 59:91-103
12 Kuhn, B., Gerber, P., Schultz-Gash, T., Stahl, M. (2005) J. Med. Chem. 48:4040-4048
13 Pearlman, D. A. (2005) J. Med. Chem. 48:7796-7807
14 Woo, H. J., Roux, B. (2005). Proc. Natl. Acad. Sci. U.S.A. 102:6825-6830
15 Sharp, K. A.; Honig, B. (1990) Annu. Rev. Biophys. Biophys. Chem. 19:301-332
16 Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson (1990), T. J. Am. Chem. Soc. 112:6127-6129
17 Foloppe, N., Hubbard, R. (2006) Curr. Med. Chem. 13:3583-3608
18 Swanson, J. M. J., Henchman, R. H., McCammon, J. A. (2004) Biophys. J. 86:67-74
19 Gohlke, H., Case, D. A. (2004) J. Comput. Chem. 25:238-250
20 Stoica, I., Kashif Sadiq, S., Coveney, P. V. (2008) J. Am. Chem. Soc. 130:2639-2648
21 A. Weis, K. Katebzadeh, P. Söderhjelm, I. Nilsson, U. Ryde (2006) J. Med. Chem., 49, 6596-6606.
22 Weber, P. C., Ohlendorf, D. H., Wendolowski, J. J., Salemme, F. R. (1989) Science 243:85-88
23 Weber, P. C., Wendolowski, J. J., Pantoliano, M. W., Salemme, F. R. (1992) J. Am. Chem. Soc. 14:3197-3200
24 Pugliese, L., Coda, A., Malcovati, M., Bolognesi, M. (1993) J. Mol. Biol. 231:698-710
25 Livnah, O., Bayer, E. A., Wilchek, M., Sussman, J. L. (1993) Prot. Natl. Acad. Sci. U.S.A. 90:5076-5080
26 Green, N. M. (1966) Biochem. J. 101:774-790
27 Green, N. M. (1975) Adv. Protein. Chem. 29:85-133
28 Green, N. M. (1990) Methods Enzymol. 184:51-67
29 Miyamoto, S., Kollman, P. A. (1993) Proteins: Struct. Funct. Genet. 16:226-245
30 Wang, J. Dixon, R. Kollman P. A. (1999) 34:69-81
31 Kuhn, B., Kollman, P. A. (2000) J. Med. Chem. 43:3786-3791
32 Brown, S. P. Mychmore, S. W. (2006) J. Chem. Inf. Model. 46:1493-1503
33 Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C. (2006) Proteins: Struct., Funct. Bioinform. 65:712-725
34 Wang, J. Cieplak, P., Kollman, P. A. (2000) J. Comput. Chem. 21:1074-1074
35 C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, J. Phys. Chem. 1993, 97, 10269-10280.
36 Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G., Head-Gordon, T. (2004) J. Chem. Phys. 120:9665-9678
37 Showalter, S. A.; Brüschweiler, R. J. Chem. Theory Comput. 2007- 3- 961-975.
38 Koller, A. N.; Schwalbe, H.; Gohlke, H. Biophys. J. Biophys. Lett. 2008, 108, L04-L06.
39 Wong, V.; Case, D. A. J. Phys. Chem. B. 2008, 112, 6013-6024.
40 Sörme, P., Pascal, A., Kahl-Knutsson, B., Leffler, H., Rini, J. M., Nilsson, U. J. (2005) J. Am. Chem. Soc. 127:1737-1743
41 Wang, J. M., Wolf, R. M., Caldwell, K. W., Kollman, P. A., Case, D. A. (2004) J. Comput. Chem. 25:1157-1174
42 B. H. Besler, K. M. Merz, P. A. Kollman, J. Comput. Chem. 1990, 11, 431-439.
43 D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.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.
44 Ryckaert, J. P, Ciccotti, G., Berendsen, H. J. C. (1977) J. Comput. Phys. 23:327-341
45 Wu, X., Brooks, B.R. (2003) Chem. Phys. Lett. 381:512-518
46 Darden, T., York, D., Pedersen, L. (1993) J. Chem. Phys. 98:10089-10092
47 D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R.

Luo, K. M. Merz, B. Wang, D. A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui. P. Beroza, C. Schafmeister, J. W. Caldwell, W. S. Ross, and P. A Kollman (2004) AMBER 8, University of California, San Francisco.
48 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.
49 J. Kongsted, U. Ryde, J. (2009) Comp.-Aided Mol. Des., 23:63-71.
50 Bishop, M., Frinks, S. (1987) J. Chem. Phys. 87:3675-3686
51 Yang, W., Bitetti-Putzer, R., Karplus, M. (2004) J. Chem. Phys. 120:2618-2628
52 Fujitani H.; Tanida, Y.; Ito, M.; Jayachandran G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. (2005) J. Chem. Phys., 123:0804108
53. M. H. M. Olsson, G. Hong, A. Warshel, J. Am. Chem. Soc. 2003. 125. 5025-5039.
54 Mongan, J.; Case, D. A.; McCammon, J. A.; *J. Comput. Chem.***2004**, *20*, 2038-2048.
55 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.
56 Nakanishi, I.; Fedorov, D. G.; Kitaura, K. Proteins, Struct. Funct. Bioinf. 2007, 68, 145-158.
57 Kongsted, J.; Söderhjelm, P.; Ryde, U., J. Phys. Chem. B., 2009, submitted.
58 Kongsted, J.; Söderhjelm, P.; Ryde, U., J. Comput.-Aided Mol. Design, submitted.

Table 1. MM/GBSA results (in kJ/mol) for the binding of biotin to avidin, using one 10 ns long simulation and a sampling frequency of 5 ps.

Subunit	ΔG_{bind}	Standard error
A	-120.9	0.5
B	-110.3	0.5
C	-119.5	0.5
D	-118.2	0.5

Table 2. MM/GBSA results (in kJ/mol) for the binding of biotin to avidin, using 20 independent simulations (200 ps long).

Subunit	ΔG_{bind}	Standard error
A	-114.0	2.1
B	-111.1	2.9
C	-117.2	2.1
D	-116.1	1.5
Plain average	-114.6	1.1

Table 3. The efficiency of energy calculations within the short simulations for the biotin–avidin complex. The number of energy calculations is increased from 1 to 41 (corresponding to a length of the production calculation of 0 to 200 ps. The actual standard deviation (calculated for 20 short simulations and four subunits) is compared to that calculated from the standard deviation of the calculation with only one energy calculation divided by the square root of the number of energy calculations. Efficiency is the quotient of those two estimates. Finally, the $s_{simu}\sqrt{CPU_{simu}}$ estimate of efficiency (see the text), is also given.

# Energy	Standard deviation		Efficiency	$s_{simu}\sqrt{CPU_{simu}}$
calculations	Actual	Estimated		
1	25.6	25.6	1.00	125.3
2	20.5	14.5	0.71	117.4
3	17.4	10.0	0.58	112.2
4	14.6	7.3	0.50	103.9
5	13.3	6.0	0.45	102.5
6	12.3	5.0	0.41	101.2
7	11.6	4.4	0.38	101.7
9	10.7	3.6	0.33	103.7
11	10.5	3.2	0.30	111.4
16	10.5	2.6	0.25	130.7
21	10.0	2.2	0.22	140.7
31	9.8	1.8	0.18	167.0
41	9.6	1.5	0.16	185.8

Table 4. The estimated binding energy (kJ/mol) for the seven biotin analogues in Figure 1 to avidin, calculated with the MM/GBSA method and our suggested protocol (the error estimate is the standard error of the average).

Ligand	#	ΔG_{bind}		Difference
	simulations	calculated	experimental	
Btn1	25	-114.4±1.0	-85.4	-29.0
Btn2	30	-102.6±1.0	-59.8	-42.8
Btn3	20	-102.2±0.8	-58.6	-43.6
Btn4	50	-112.2±0.8	-36.8	-75.4
Btn5	40	-67.6±1.0	-34.3	-33.3
Btn6	20	-64.5±0.7	-20.9	-43.6
Btn7	20	-19.7±0.5	-18.8	-0.9

Figure 1. The seven biotin analogues used in this study. a) Btn1 (biotin), b) – g) Btn2–Btn7.

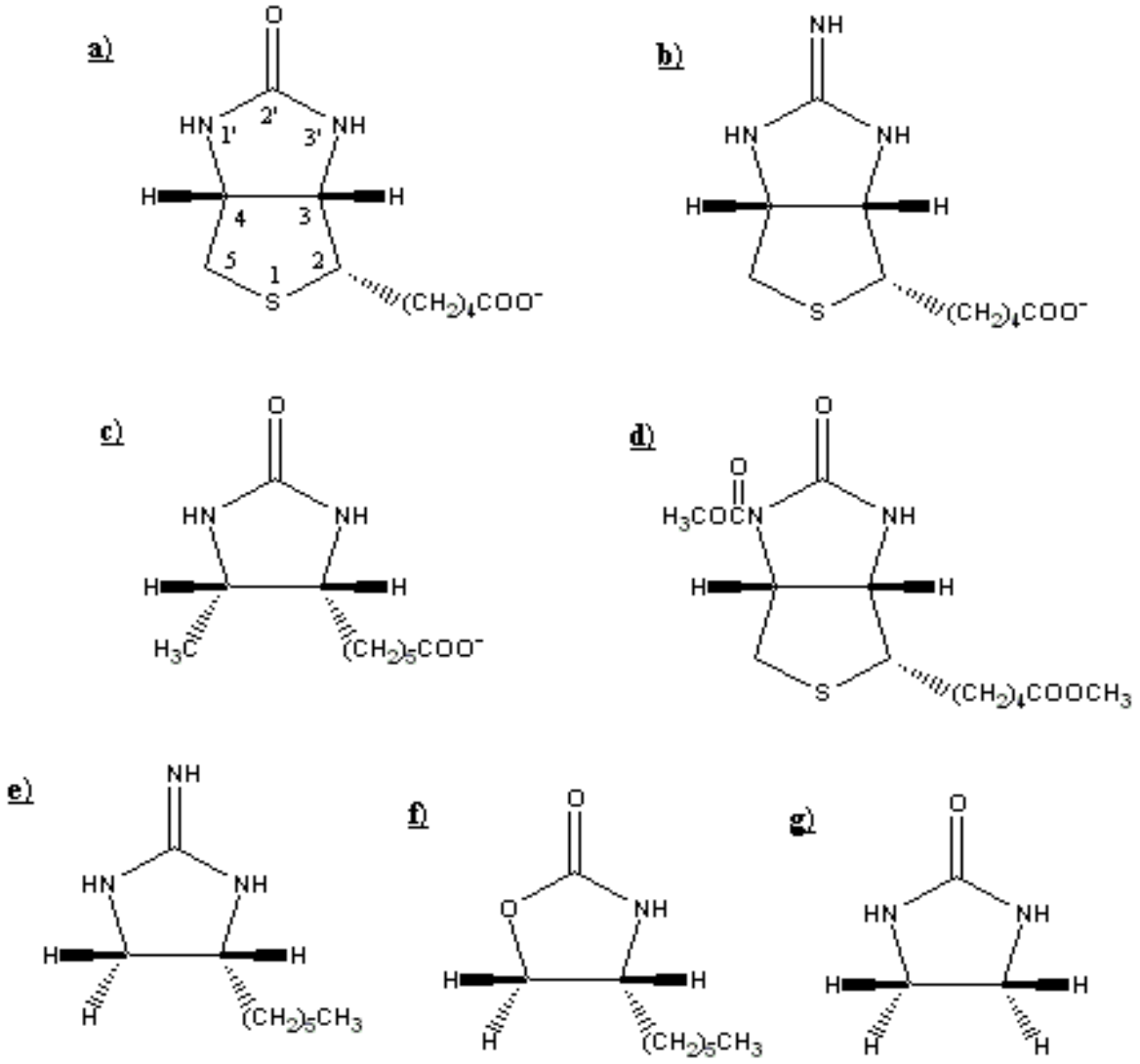

Figure 2. The 80 MM/GBSA estimates for the biotin–avidin complex, using 20 independent simulations.

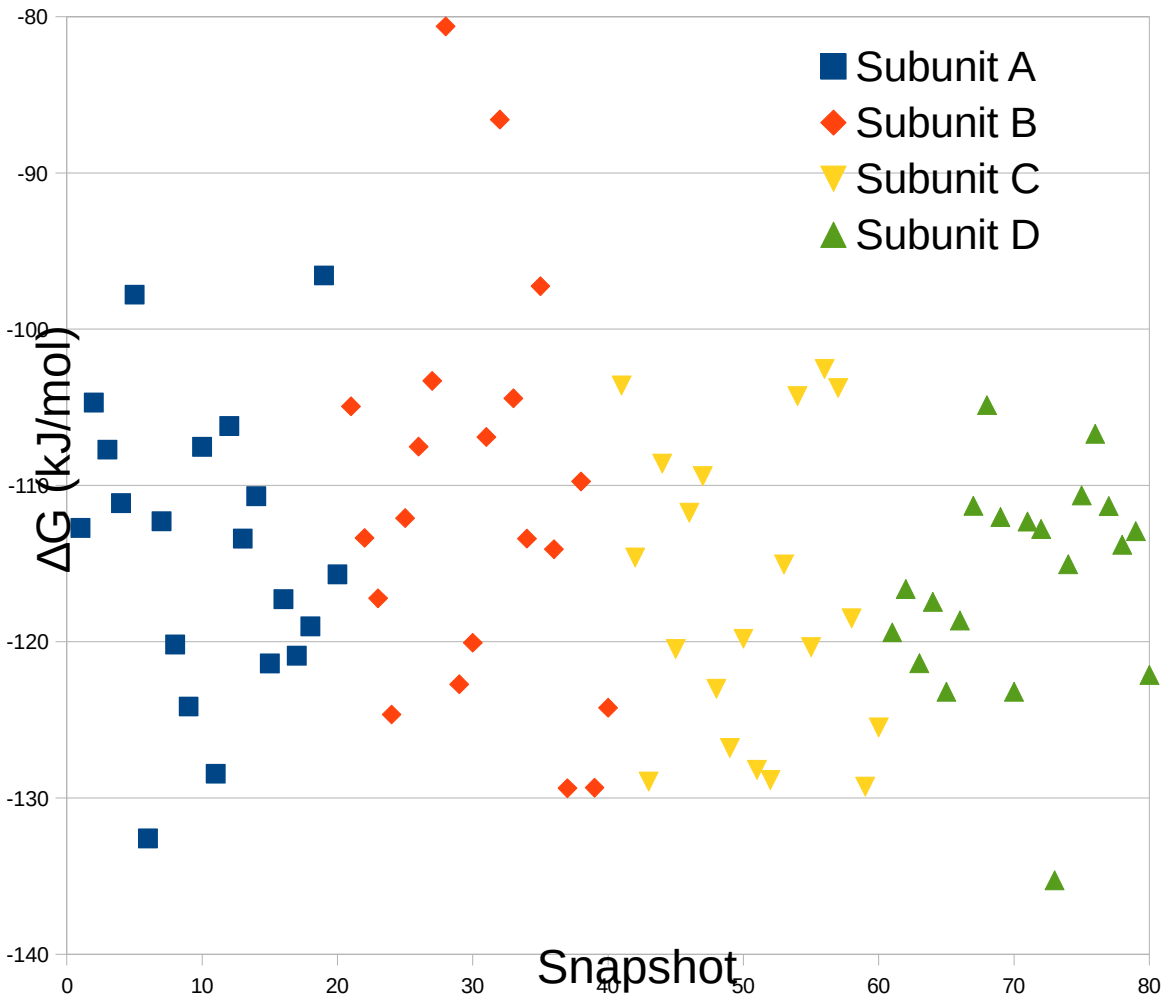

Figure 3. Distribution of the correlation time of the MM/GBSA estimates for the biotin–avidin complex. The cumulative frequency of the correlation times is shown, using either all the 400 ps simulation or only the last 200 ps. The plots are based on data from four subunits in 20 separate 400 ps simulations.

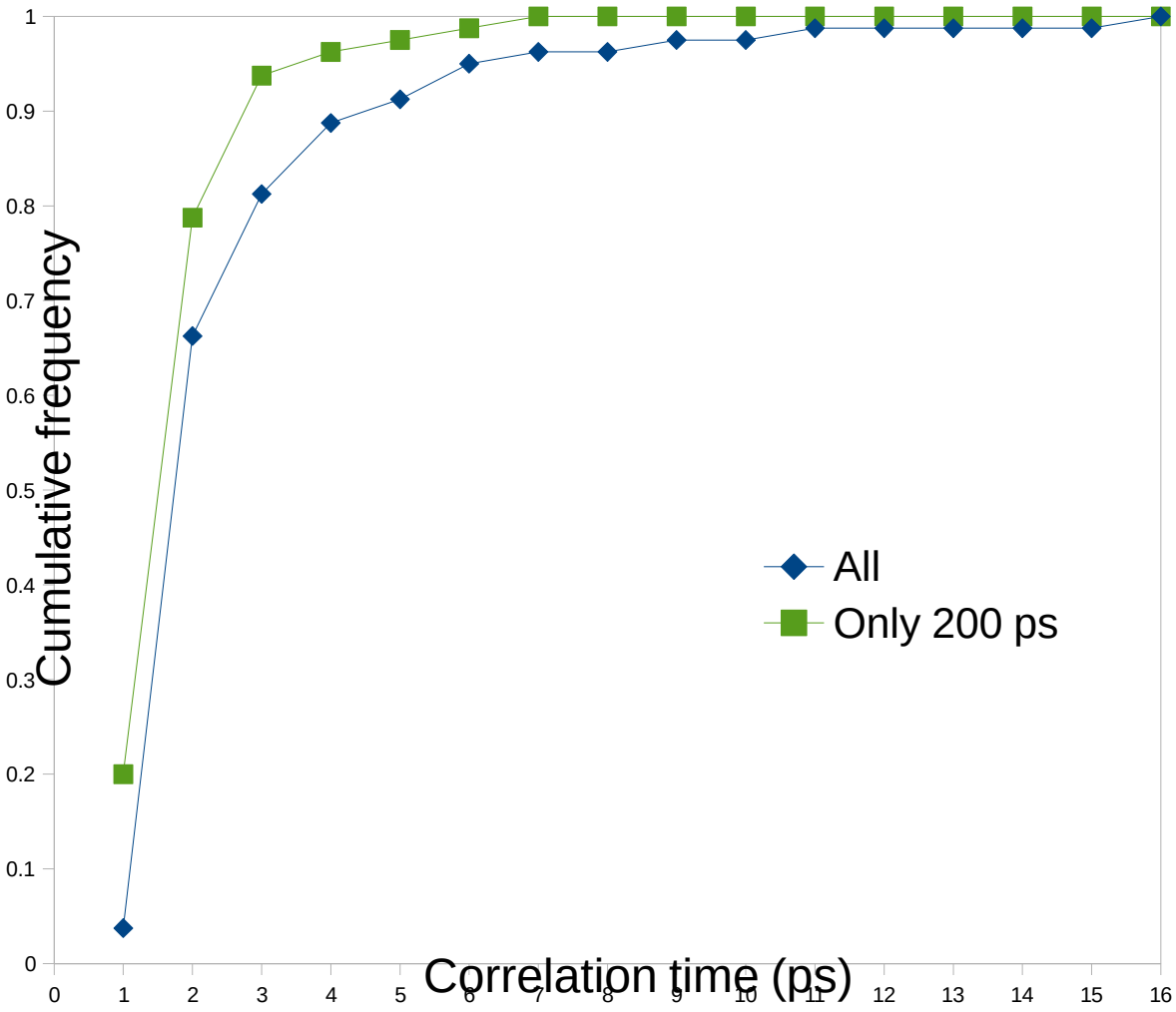

Figure 4. The variation of the MM/GBSA ΔG_{bind} energy with the length of the equilibration time for the biotin–avidin complex. Several different lengths of the production simulation time were tested. Full means that the rest of the 400 ps simulation was used.

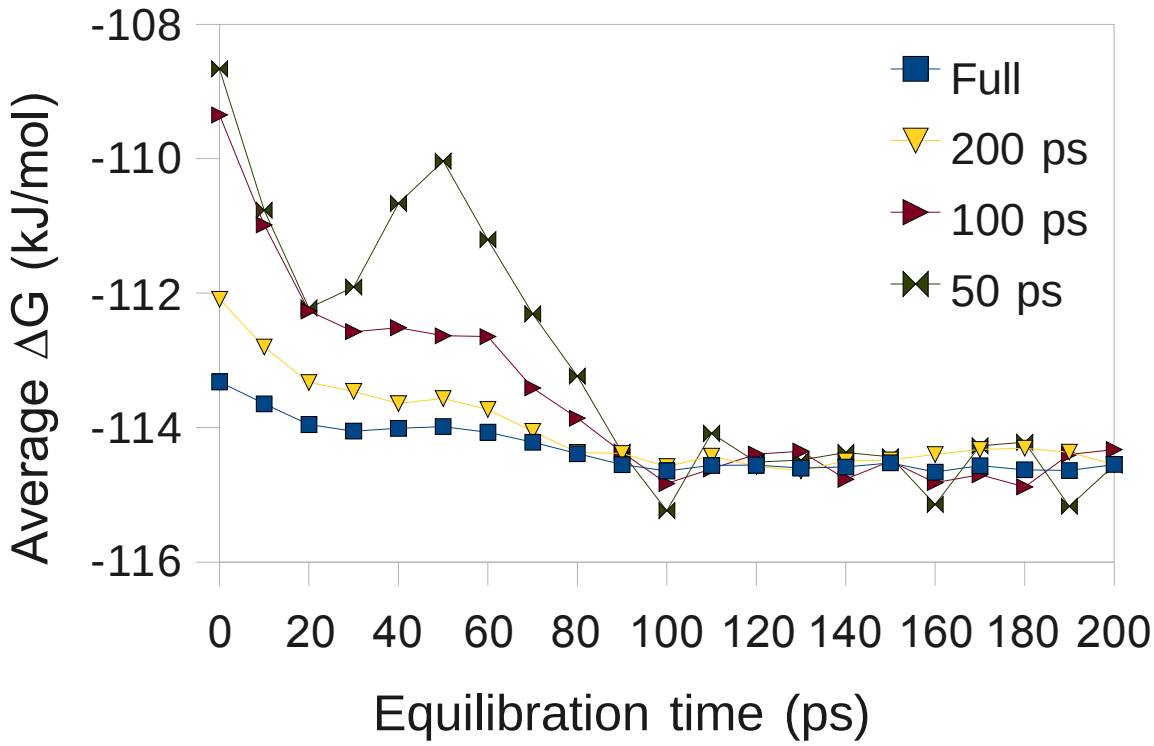

Figure 5. The variation of the MM/GBSA ΔG_{bind} energy with the length of the production simulation for the biotin–avidin complex. Several different lengths of the equilibration time (0–200 ps) were tested.

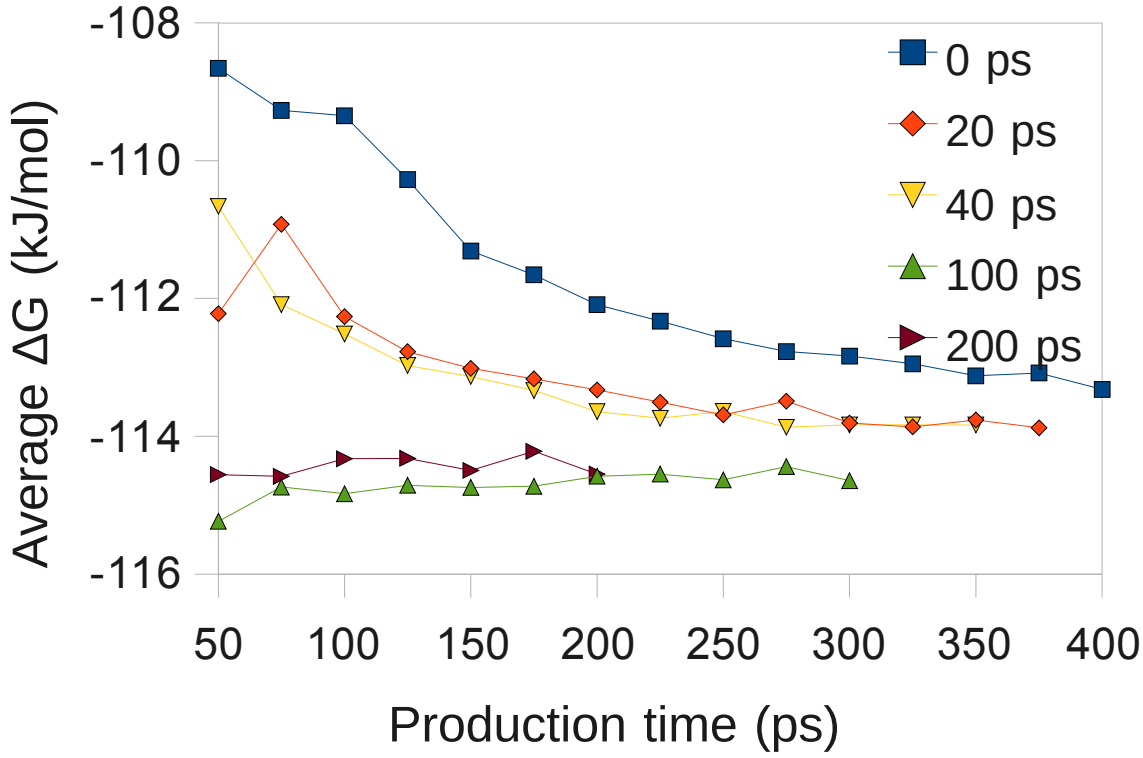

Figure 6. The variation of the standard error of the ΔG_{bind} energy with the length of the production simulation for the biotin–avidin complex. Several different lengths of the equilibration time (0–200 ps) were tested.

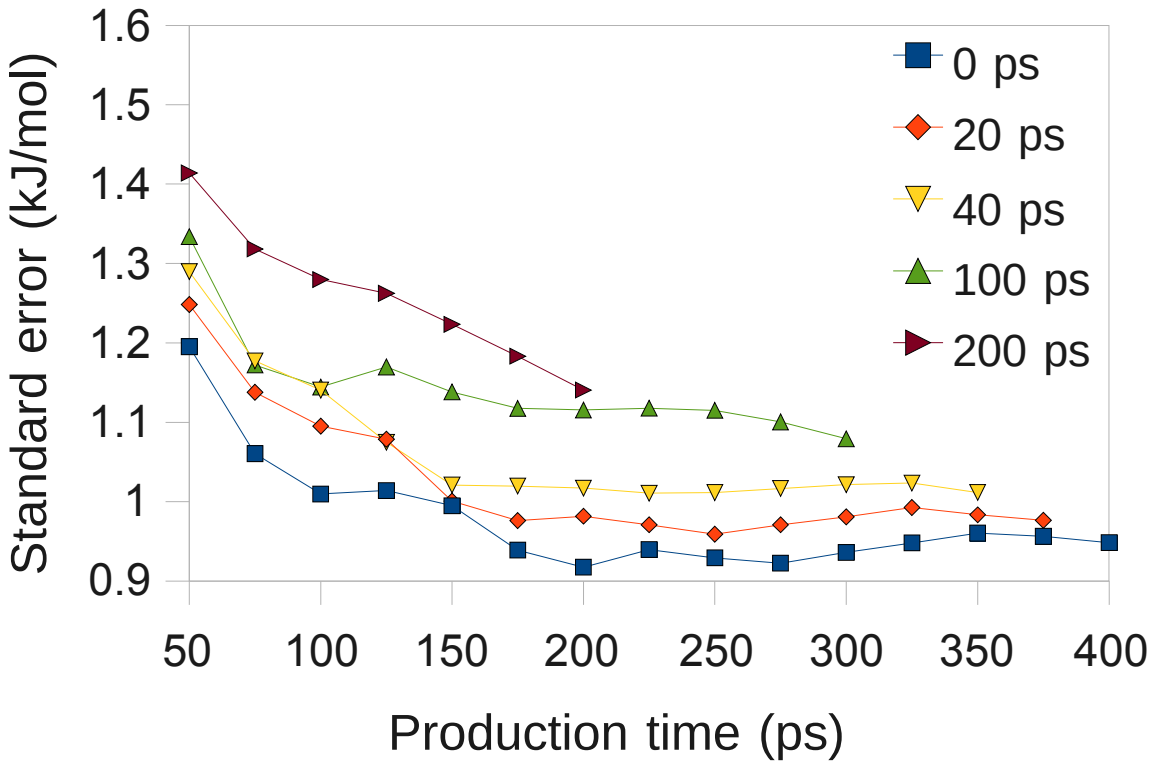

Figure 7. Correlation between experimental and calculated binding energies for the seven biotin analogues to avidin. The upper line represents the perfect correlation, whereas the lower line represents the former line, translated by the signed average error (–39.4 kJ/mol).