Binding affinities by alchemical perturbation using QM/MM with a large QM system and polarizable MM model.

Genheden, Samuel; Ryde, Ulf; Söderhjelm, Pär

Published in:
Journal of Computational Chemistry

DOI:
10.1002/jcc.24048

2015

Binding affinities by alchemical perturbation using QM/MM with a large QM system and polarizable MM model

Samuel Genheden [a], Ulf Ryde [b], Pär Söderhjelm [c]*

[a] School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

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

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

Correspondence to Pär Söderhjelm, E-mail: Par.Soderhjelm@bpc.lu.se,

Tel: +46 – 46 2228161, Fax: +46 – 46 2228648

2016-01-06

Abstract

The most general way to improve the accuracy of binding-affinity calculations for protein–ligand systems is to use quantum-mechanical (QM) methods together with rigorous alchemical-perturbation methods. We explore this approach by calculating the relative binding free energy of two synthetic disaccharides binding to galectin-3 at a reasonably high QM level (dispersion-corrected DFT with a triple-zeta basis set) and with a sufficiently large QM system to include all short-range interactions with the ligand (744–748 atoms). The rest of the protein is treated as a collection of atomic multipoles (up to quadrupoles) and polarizabilities. Several methods for evaluating the binding free energy from the 3600 QM calculations are investigated in terms of stability and accuracy. In particular, methods using QM calculations only at the endpoints of the transformation are compared with the recently proposed *non-Boltzmann Bennett acceptance ratio* (NBB) method that uses QM calculations at several stages of the transformation. Unfortunately, none of the rigorous approaches give sufficient statistical precision. However, a novel approximate method, involving the direct use of QM energies in the *Bennett acceptance ratio* method, gives similar results as NBB but with better precision, ~3 kJ/mol. The statistical error can be further reduced by performing a greater number of QM calculations.

Key words: Binding affinity, quantum chemistry, dispersion-corrected DFT, protein–ligand interaction, free-energy perturbation, Bennett acceptance ratio, alchemical transformation, QM/MM, galectin-3.

Introduction

The binding of a small ligand to a protein depends on a delicate balance between protein–ligand, protein–water, and ligand–water interactions, as well as another balance between energetic and entropic contributions to the binding. An accurate estimation of the binding free energy therefore requires accurate descriptions of these interactions, but also a thorough account of the distribution of structures dictated by these interactions. The latter is typically achieved through molecular dynamics (MD) or Monte Carlo simulations, which generate ensembles of molecular configurations representative for the system of interest.

Owing to error cancellation between the various interactions, it has turned out to be easier to compute the *difference* in binding free energy between two ligands binding to the same protein, than to compute the *absolute* binding free energy for a given system [1]. For the practical aspect of predicting binding free energies of unknown systems, for example in the pharmaceutical industry, such *relative* free energies are also the most relevant ones [2], as one is most often interested in finding a set of drug candidates for a given protein target and only rarely in comparing different targets with each other.

One of the most rigorous methods for computing relative binding free energies at the molecular mechanics (MM) level is *alchemical perturbation (*AP). This method typically involves the sampling of configurations along an alchemical transformation of one ligand into the other while bound to the protein and free in solution, and the subsequent evaluation of the free-energy differences through exponential averaging [3], thermodynamic integration [4], or the Bennett acceptance ratio (BAR) method [5], of which the latter usually shows the best convergence properties [6].

The reliability of the AP approach typically depends on how similar the two ligands are, but also on their size [7]. Clearly, if the ligands are very different, one may expect that they have different *binding modes*, e.g. differences in ligand orientation, ligand conformation, receptor structure, or presence of water molecules in the binding site, and thus adequate sampling of both states becomes problematic [2]. However, in other cases, the two ligands can be very similar, but the alchemical approach still fails to reproduce the experimental difference in binding free energy [8, 9, 10]. In such cases, the MM force field may not be accurate enough to describe the interactions [11]. Indeed, a recent study of the sensitivity on force field parameters found differences in binding free energies of several kcal/mol when the parameters were varied in a realistic range [12].

There are several ways to improve the force field used in binding free energy estimations, e.g. performing a more careful parametrization of the ligand [13] or employing a force field that explicitly includes electronic polarization [14, 15]. The most general approach, however, is to use quantum-mechanical (QM) calculations, as these are able to treat all systems in an unbiased manner and can be systematically improved [16]. Unfortunately, a level of theory that would guarantee high accuracy, such as coupled cluster with singles, doubles and perturbative triples (CCSD(T)), is very computationally expensive and usually beyond reach for protein–ligand systems, although such calculations based on fragmentation are emerging [17, 18]. Even second-order Møller–Plesset (MP2) calculations are relatively rare [19, 20].

On the other hand, several protein–ligand calculations with density functional theory (DFT) have been published, either using a QM-cluster approach [21], a combined QM and MM approach (QM/MM) [22], or a fractionation approach [23]. In the QM/MM calculations, the QM system has sometimes constituted only the ligand, so that the protein–ligand interactions are modelled by MM,

whereas some studies also include a part of the protein in the QM system to describe the interactions at the DFT level. In the latter case, it is essential that a dispersion-corrected DFT variant and a sufficiently large basis set are used [24, 25, 26].

Only a few of the previous QM studies of ligand binding have employed an AP approach [27, 28], despite that this is the gold standard in binding free-energy calculations at the MM level. Most other studies used either geometries from docking [29], QM/MM-optimised geometries [24, 25], average structures from molecular dynamics [30], or a limited ensemble of structures from molecular dynamics [31, 32]. In our experience, the approximations introduced by not using an AP approach are so severe that they typically hide the gain of switching from MM to QM [16, 32, 33, 34, 35].

The direct application of QM in alchemical AP calculations would require extensive sampling at the QM level, which is not feasible for a large system. One way to get around this problem is to perform the sampling at a more approximate reference level (typically MM) and then include the free energy of changing the model from the reference to QM (or QM/MM) [36, 37, 38, 39]. In principle, the corrections need to be evaluated only at the endpoints of the transformation, thus leading to methods like QM/FEP [40, 41], ABC-FEP [42], QTCP [43], paradynamics [44], and related methods [45]. These methods have mainly been used to calculate solvation free energies [46] or reaction free energies in enzymes, but also binding free energies [45]. A reasonably good overlap between the phase space of the reference and QM potentials is required to obtain convergent results. This can be achieved by improving the reference potential [47, 48, 49], choosing the most compatible reference potential [50] or by introducing intermediate states [44]. Another, related method, uses a Metropolis–Hastings algorithm to enable sampling at the QM level but employs a reference potential to reduce the number of QM energy evaluations [51]. With this approach, statistical errors of ~1 kJ/mol on binding free energies could be achieved, because the small QM system permitted 20,000 QM evaluations per λ value [52].

Another way of approaching this problem, which will be investigated in the current paper, is to use the free-energy expressions (e.g. BAR) directly, but to evaluate the interaction energies by QM instead of MM. This is similar in spirit to some methods used for enzyme reactions [53, 54], but also including sampling of the QM system at the MM level. Although this is not rigorously correct, because it assumes that the ensembles generated by MM are identical to those that would be generated by QM, one may expect that the major effects of using QM will be captured.

Recently, a method was proposed that enables the rigorous computation of free energies based on ensembles generated with a biased potential [55]. The method, denoted n*on-Boltzmann Bennett acceptance ratio* (NBB), uses the knowledge of the bias at each snapshot to reweight the free-energy contributions so that the free energy for the unbiased potential is obtained. Applied to the current problem, the bias is the difference between the MM energy and the QM energy for the same geometry, and the result of the NBB calculation is the desired free energy at the QM level. This type of reweighting has previously been shown to give improved results over the endpoint approach in the calculation of hydration free energies [56, 57], but the results for binding free energies have been inconclusive [58].

The requirement to use a reasonably accurate QM method restricts the system size that can be treated by QM to a few hundred atoms. One therefore expects the results to depend on how the rest of the system is treated. Most previous QM/MM studies have modelled the surroundings (protein or water) as point charges interacting with the QM charge density of the inner system, whereas the van der Waals interactions have been modelled by a pure MM potential [59]. However, test calculations have pointed out that there may be a dependence on the model choice that extends to large distances

[19] and that therefore a more accurate account of the electrostatics is advantageous. One such model, which we denote *polarizable multipole interaction* (PMI), uses atomic multipoles and polarizabilities to describe the surrounding protein, whereas a standard MM potential is retained for the van der Waals interactions, whose model dependence has a shorter range [19]. This method has previously been applied together with a fragmented QM method, the PMISP approach [19, 60] and similar methods have been developed by other authors [61, 62, 63].

In this study, we perform a large set of QM calculations, supplemented by a PMI model of the surroundings, on a series of ensembles generated by AP. We compare various approaches to use the calculated energies to compute binding free energies. In particular, we investigate the stability of the endpoint and NBB approaches, as well as various approximations that use the QM energies in a more efficient way so that convergent results can be obtained. Moreover, we discuss how to take the surroundings into account in an accurate manner. As a test case, we study the binding of two synthetic disaccharide ligands to galectin-3 with a relative binding free energy of 10 kJ/mol [64].

Methods

Free energy methods

To compute the relative binding free energy $\Delta\Delta G$ of two ligands, **L** and **L'**, we apply the thermodynamic cycle in Figure 1, which leads to

$$\Delta\Delta G = \Delta G_{L'} - \Delta G_{L} = \Delta G_{bound} - \Delta G_{free}. \quad (1)$$

where ΔG_{bound} and ΔG_{free} are the free energies of transforming **L** to **L'** when they are bound to the protein or free in solution, respectively. The rest of this section deals with the computation of ΔG_{bound}. Exactly the same methods can then be applied to compute ΔG_{free}.

At the MM level, ΔG_{bound} is computed using the Bennett acceptance ratio (BAR) method [5]. In this method, a finite number of λ values are selected between 0 and 1, and for each λ, an MD simulation is run with the potential

$$V_\lambda = (1 - \lambda)V_0 + \lambda V_1, \quad (2)$$

where V_0 is the potential with **L** and V_1 is the potential with **L'** (due to the soft-core potentials [65,66], V_0 and V_1 also depend on λ, as detailed in the Supporting information). Thus, the scaling factor λ transforms the potential from that of **L** (λ = 0) to that of **L'** (λ = 1). For each neighbouring pair of λ values, *A* and *B*, the free energy difference between the two states can be written as

$$\Delta G^{A\to B} = kT\left(\ln\frac{\langle f(V_A - V_B + C)\rangle_B}{\langle f(V_B - V_A + C)\rangle_A}\right) + C \quad (3)$$

where $f(x) = (1 + \exp(x/kT))^{-1}$ is the Fermi function, k is the Boltzmann constant, T is the temperature, and C is a constant. An iterative procedure is then applied, in which one looks for a value of C that makes the first term on the right-hand side of Eq. 3 vanish. Finally, the total ΔG_{bound} is obtained by summing over all λ intervals.

At the QM level, a straightforward application of BAR can normally not be done, as it would

require MD simulations with the QM potential, which is computationally too demanding. We use two types of methods to get around this problem and still retain the rigorousness. The first type, illustrated in Figure 2a, is an endpoint approach, previously applied in methods such as QM/FEP, QTCP, and paradynamics [40–45]. Here, a thermodynamic cycle is used to obtain

$$\Delta G_{bound}^{QM} = \Delta G_{bound}^{MM} + \Delta G_{L'}^{MM \to QM} - \Delta G_{L}^{MM \to QM} \tag{4}$$

where ΔG_{bound}^{MM} is the free energy of the whole transformation at the MM level, obtained by the standard BAR approach for the various subintervals, and the remaining terms are correction terms for going from the MM potential to the QM potential. Note that these have to be evaluated only at the endpoints of the transformation, i.e. for each of the ligands in its own ensemble. Each correction term is either rigorously evaluated using exponential averaging:

$$\begin{aligned} \Delta G_{L}^{MM \to QM} &= -kT \ln \left\langle \exp\left(-\left[V_{L}^{QM} - V_{L}^{MM} \right] / kT \right) \right\rangle_{L} \\ \Delta G_{L'}^{MM \to QM} &= -kT \ln \left\langle \exp\left(-\left[V_{L'}^{QM} - V_{L'}^{MM} \right] / kT \right) \right\rangle_{L'} \end{aligned} \tag{5}$$

or approximated by a plain average (corresponding to the limit of infinite temperature):

$$\begin{aligned} \Delta G_{L,\,plain}^{MM \to QM} &= \left\langle V_{L}^{QM} - V_{L}^{MM} \right\rangle_{L} \\ \Delta G_{L',\,plain}^{MM \to QM} &= \left\langle V_{L'}^{QM} - V_{L'}^{MM} \right\rangle_{L'} \end{aligned} \tag{6}$$

To reduce the statistical noise, we also tested the additional approximation to compute these corrections over the same ensemble of configurations:

$$\begin{aligned} \Delta G_{L,\,plain,\,\lambda}^{MM \to QM} &= \left\langle V_{L}^{QM} - V_{L}^{MM} \right\rangle_{\lambda} \\ \Delta G_{L',\,plain,\,\lambda}^{MM \to QM} &= \left\langle V_{L'}^{QM} - V_{L'}^{MM} \right\rangle_{\lambda} \end{aligned} \tag{7}$$

where λ denotes the particular λ value with which the ensemble was generated.

The second type of method to calculate a QM free energy is illustrated in Figure 2b. In this approach, the QM free energy is obtained for each subinterval A→B of the transformation, using the same iterative approach as in BAR. However, the ensemble available for computing the QM energies has been obtained with an MM-based potential; it is as if the ensemble has been obtained in a *biased* simulation, with the bias corresponding to the difference between the approximate (MM) and true (QM) potentials. This situation can be resolved by the non-Boltzmann Bennett acceptance ratio (NBB) method [55]. In this method, when taking the bias into account, the BAR expression has to be modified into

$$\Delta G^{A \to B} = kT \left(\ln \frac{\left\langle f\left(V_{A}^{QM} - V_{B}^{QM} + C \right) \exp\left(V_{B}^{bias} / kT \right) \right\rangle_{B} \left\langle \exp\left(V_{A}^{bias} / kT \right) \right\rangle_{A}}{\left\langle f\left(V_{B}^{QM} - V_{A}^{QM} + C \right) \exp\left(V_{A}^{bias} / kT \right) \right\rangle_{A} \left\langle \exp\left(V_{B}^{bias} / kT \right) \right\rangle_{B}} \right) + C \tag{8}$$

where $V_{\lambda}^{bias} = V_{\lambda}^{MM} - V_{\lambda}^{QM}$. In this equation, V_{λ}^{QM} is computed as a linear combination of QM energies for the two ligands, according to Eq. 2. Thus, this method requires QM and MM evaluations of each of the ligands at each of the snapshots from each of the simulations with

different λ values. Some recent studies have employed an *indirect* version of the NBB method [56,57], in which the MM energy function is used as the true potential for all intermediate λ values (as illustrated by the dashed line in Fig. 2b and explained in detail in the Supporting information). In principle, the two versions should give the same result, because the free energy is independent of the transformation path. However, they might have different convergence properties, so we tested both versions in this study. If there are more than two intermediate λ values, the indirect version requires fewer QM calculations [56, 57].

Owing to the difficulty of obtaining statistically converged results with the NBB method, we also tested to neglect the fact that the ensembles were biased, i.e. to set $V_A^{bias} = V_B^{bias} = 0$ in Eq. 8, thus recovering the BAR expression but with QM energies instead of MM energies. This method will simply be denoted BAR in the following. The BAR and NBB calculations were performed with a modified version of the *PYMBAR* script [67]. The statistical uncertainties were estimated using bootstrapping (sampling with replacement; 100 times) or, for plain averages, as the standard deviation divided by the square root of the sample size.

System

We studied the relative binding free energy of the two ligands shown in Figure 3 to the carbohydrate-recognition domain of galectin-3 (gal3). The ligands will be denoted **3** and **4** to be consistent with the original publication [64]. The crystal structure of the complex gal3–**3** was used as the starting structure for the calculations. The structure of gal3–**4** was obtained by replacing the methoxy group by a fluorine atom. The preparation of the protein has been detailed previously [68]. All Asp and Glu residues were negatively charged, and all Lys and Arg residues were positively charged. The histidine at the binding site (H158) was protonated on the ND1 atom, and the other three histidine residues were protonated on the NE2 atom, in accordance with previous studies and NMR measurements [69]. The protein was described with the Amber99SB force field [70], whereas the ligands were described with the general Amber force field [71] with charges obtained by the restrained electrostatic potential (RESP) method [72]. The protein–ligand complexes or the free ligands were solvated in a pre-equilibrated truncated octahedral box of TIP3P water molecules [73] extending at least 10 Å from the solute.

For the QM calculations, we define a subsystem S_1, which includes the ligand and all protein atoms or water molecules within 6 Å of it. This system is shown in Figure 4 for the bound state and it contains 748 or 744 atoms depending on the ligand. For the free state it includes only the ligand and water molecules (527 or 523 atoms in total). The selection of protein atoms was identical for all snapshots and was based on the last snapshot of the simulation at $\lambda = 0.01$ for the larger ligand **3**. We verified that none of the protein atoms outside S_1 ever came closer than 5 Å to the ligand; more details on the movements are given in the Supporting information (Fig. S1). The selection was based on the bond topology of the system and did not necessarily include full residues; instead, a minimal system was selected, containing all atoms within 6 Å as well as atoms needed to avoid cutting through polar or double bonds (thus giving chemically reasonable groups such as full aromatic rings, $CH_3NHCOCH_3$ back-bone groups, etc.; cf. Figure 4). Dangling bonds were capped by hydrogen atoms at standard distances. In the MM calculations of S_1, the charges on the atoms close to the truncation were not modified, as earlier investigations have shown a negligible impact of such corrections on relative interaction energies, provided that the truncations are sufficiently far from the transformed atoms [74].

For the PMI calculations, we also define a bigger subsystem S_2 that consists of the ligand, all

protein atoms (for the bound state), and all water molecules inside an adjustable cut-off radius, *R*, from the ligand.

Simulations

The total potential energy of the system was a linear combination (Eq. 2) of the energies for the system with ligand **3** (V_0) and for the system with ligand **4** (V_1), either bound to the protein or free in solution. The simulations were performed at eleven values of λ, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 0.99 (the endpoints λ = 0 and λ = 1 cannot be handled with the version of Amber that we used). Soft-core versions of the Lennard-Jones and Coulomb potentials were used to ensure that van der Waals and electrostatic forces could be scaled simultaneously [65, 66]. Both ligands were simulated simultaneously using a dual topology.

All MD simulations were run by the *sander* module of Amber 11 [75]. The integration time step was 2 fs and SHAKE constraints [76] were applied to all bonds involving hydrogen atoms. The use of a large time step and SHAKE constraints has previously been found to introduce errors of ~2 kJ/mol when the ensemble is used for QM corrections of the total energy [56], but we expect the effect on interaction energies to be much smaller. The Lennard-Jones potential was truncated at 8 Å and a long-range correction was obtained with a continuum approach [77]. Electrostatics were treated with particle-mesh Ewald summation [78] using a real-space cut-off of 8 Å, a fourth-order B-spline interpolation, and a tolerance of 10^{-5}. The non-bonded pair list was updated every 50 fs. The temperature was kept at 300 K using Langevin dynamics [79] with a collision frequency of 2 ps^{-1}, and the pressure was kept at 1 atm using an isotropic weak-coupling algorithm [80] with a relaxation time of 1 ps.

For each value of λ, the following simulation protocol was applied: The system was first minimised by 500 steps of steepest descent, using a harmonic restraint towards the crystal structure with a force constant of 418.4 kJ/mol/$Å^2$ on all atoms, except water molecules and hydrogen atoms, followed by a 20 ps MD simulation in the NPT ensemble with the same restraints but with the force constant halved, and a 500 ps MD simulation in the same ensemble without any restraints. Finally, a 1000 ps NPT simulation without restraints was performed, in which structures and energies were saved every 10 ps.

QM energies

Using the Turbomole quantum-chemical package [81, 82], we calculated the QM energy of S_1 for all 100 snapshots at λ = 0.01, 0.30, and 0.99. For each snapshot, two QM interaction energies were computed, one with each ligand. The coordinates of all atoms, except those directly involved in the transformation, were identical in these two calculations, and the coordinates of all atoms were obtained directly from the MD program.

The BLYP functional [83, 84] with the def2-SV(P) basis set [85] was used, together with the DFT-D3 dispersion correction [86]. This functional has been shown to reproduce high-level QM interaction energies within ~3 kJ/mol [87] , and within 1 kJ/mol for a fluoromethane–water interaction relevant for the current study (see Figure S2 in the Supporting information). The same functional has previously been used for QM corrections of binding free energies [50, 52]. To speed up the calculation, the resolution-of-identity approximation with the corresponding auxiliary basis set [88, 89] was used. We used a finer-than-default integration grid, m4. Only rigid interaction energies were calculated, i.e. the same geometry was assumed for the complex, ligand, and ligand-

free complex. Counterpoise corrections were calculated on every fifth snapshot [90]. For every tenth snapshot, we also calculated the energy using a larger basis set, def2-TZVP [91]. On average, the def2-SV(P) calculations took about 1 day and the def2-TZVP calculations three days, but the variation among the individual calculations was large. The dispersion correction was computed with the dftd3 software [92], with default parameters for the BLYP functional.

PMI energies

For each snapshot at λ = 0.01, 0.30, and 0.99, distributed multipoles up to quadrupoles as well as anisotropic polarizabilities were calculated with the LoProp method [93] using Molcas 7 [94, 95]. Again, the BLYP functional with the def2-SV(P) basis set was used. Multipoles and polarizabilities (hereafter denoted *properties*) were determined in the atomic nuclei and in the bond centres. The systems containing protein segments were initially cut into residue-sized fragments (not necessarily full residues), for which the properties were computed. Then, a set of properties for the whole system was constructed by combining the fragment properties through a MFCC-like procedure (molecular fractionation with conjugate caps) [60, 96].

The QM/PMI estimate of the interaction energy between the ligand and its surrounding was computed as

$$E^{QM/PMI} = E_{S1}^{QM} - E_{S1}^{PMI} + E_{S2}^{PMI} \quad (9)$$

where E_{S1}^{QM} is the QM interaction energy between the ligand and the rest of S_1, E_{S1}^{PMI} is the PMI interaction energy between the ligand and the rest of S_1, and E_{S2}^{PMI} is the PMI interaction energy between the ligand and the rest of S_2. Each of the PMI interaction energies is in turn the difference between three self-consistent PMI energies: for the complex, for the ligand, and for the complex with the ligand removed. The intramolecular polarization was treated as described previously [60].

Results and Discussion

Binding free energies at the MM level

We calculated the relative binding free energy of ligands **3** and **4** with the AP approach using the thermodynamic cycle in Figure 1. The two ligands differs in a –F group being replaced by –OMe (Figure 3), which represents a typical perturbation for an AP calculation of relative binding affinities. The results of the BAR calculations using the Amber force field are given in Table 1. The calculated relative binding free energy is 3.6±0.5 kJ/mol (ligand **4** binds worse than ligand **3**). This is 7 kJ/mol lower than the experimental result [64] and one aim of this paper is to see if we can improve this estimate by using QM calculations – it has often been reported that halogens are difficult to model by MM methods [97,98].

It is common practice to base QM corrections on interaction energies between the ligand and its surrounding instead of total energies [40,45,99,46,100]. The use of interaction energies reduce the effect of mismatches between the intramolecular structure between the MM and QM descriptions and thus leads to a better overlap between the MM and QM phase spaces [46]. However,
it also introduces approximations, primarily that the QM correction to the intramolecular energy difference between the bound and free ligand is assumed to be negligible or at least identical for the two ligands. To estimate the magnitude of these approximations for our system, we first tested the

use of interaction energies at the MM level. As shown in Table 1, the total effect of ignoring the variation in conformational energies is ~3 kJ/mol and we can correct the final result by subtracting this energy (this is equivalent to starting from the full system MM free energy and adding a MM → QM correction based on interaction energies, as done in e.g. Ref. 100). The neglect of QM corrections on the intramolecular energies can be expected to have an even smaller effect. For future comparisons, it is worth noticing from Table 1 that the statistical error is reduced when using interaction energies, even at the MM level. This is expected, because the intramolecular degrees of freedom are practically removed from the free-energy calculation, thus reducing the noise.

To reduce the number of (computationally expensive) QM calculations, we also investigated whether we could use only a subset of the λ values, as has successfully been done before [9]). Based on how the free-energy contributions depend on λ (Figure 5), we decided to use only three values: the two endpoints (λ=0.01 and λ=0.99) almost corresponding to simulation of the actual ligands **3** and **4**, respectively, and one intermediate point (λ=0.30). When this approximation was applied to the BAR calculations with total energies, it gave an error of ~2 kJ/mol. However, when it was applied to the BAR calculations with interaction energies, it gave a negligible error of only 0.3 kJ/mol, due to a cancellation between slightly larger effects on ΔG_{bound} and ΔG_{free}. Similar cancellation was obtained using interaction energies for the truncated S_1 system (cf. Table 1) and for other system sizes (see Figure S3 in the Supporting information). The overall effect on the free energy difference of truncating the system is 4 kJ/mol. Although this is a rather small effect, it still suggests that the surroundings have to be taken into account to get accurate results.

QM corrections for system S_1

For each of the 100 snapshots from each simulation, the QM interaction energy was calculated for system S_1 (containing all atoms within 6 Å of the ligand; shown in Figure 4). Relative free energies at the QM level were computed in several different ways. First, the BAR method was used without modification, i.e. simply replacing the individual MM interaction energies by the corresponding QM energies. Second, the fact that the snapshots were taken from a biased ensemble was taken into account by using the NBB method. In Table 2, the resulting QM free energies are compared to the corresponding MM free energies. The QM free energies computed with BAR have a slightly worse precision than the MM free energies (~1 kJ/mol) and give an estimate of the QM correction (difference between a QM and a MM description) of –5±1 kJ/mol. On the other hand, the more rigorous NBB estimate suffers from much larger statistical errors (5–13 kJ/mol) and is practically useless. For example, it gives a QM correction of 0±14 kJ/mol. Nevertheless, it is reassuring that for both states, the BAR and NBB results agree within the statistical error. This indicates that the BAR estimate, although based on the inaccurate MM sampling, is a reasonable estimate of the true result. The *indirect* NBB method [56,57] gives even larger statistical error than the NBB method, suggesting a possible numerical advantage of using a QM-based energy function for the intermediate λ value. Thus, the indirect version of NBB will not be further discussed.

These QM corrections to the free energies can be compared to the endpoint corrections (using the thermodynamic cycle in Figure 2a), which are also listed in Table 2. Strictly, each correction should be computed by exponential averages according to Eq. 5. However, owing to the large spread in individual QM–MM energy differences, this approach gives too large statistical errors (up to 23 kJ/mol) to be of any use. In fact, for each of the four endpoint steps, the result is completely determined by a single snapshot, which is of course not acceptable. If we instead use plain averages (Eq. 6), a total QM correction of –12±8 kJ/mol is obtained. This is an enthalpic correction, which would be equal to the free-energy correction if the QM and MM surfaces have equal shapes. If this

assumption is almost satisfied (with only a few outliers), the plain average result would be more stable with respect to insufficient sampling. However, the large statistical error indicates that the assumption is far from satisfied in this case.

An atom-wise analysis shows that a large part of the statistical uncertainty in the plain average result stems from interactions involving atoms that are present in both ligands. If those interactions are not sufficiently sampled, the corresponding QM corrections do not cancel between the ligands, despite their tendency to cancel in the limit of perfect sampling (if assuming identical binding poses for the two ligands). Therefore, we also tested the additional approximation to evaluate both endpoint corrections using only snapshots from the λ=0.01 simulation (Eq. 7). This gives a QM correction of –12±1 kJ/mol, i.e. the same net result as Eq. 6, but with a much smaller statistical error owing to a more pronounced error cancellation between the V_0 and V_1 energies when these are evaluated using the same coordinates. Almost the same results (2 kJ/mol difference) are obtained when using the λ =0.30 coordinates, whereas the λ=0.99 coordinates give unphysical short contacts for the larger ligand **3** and therefore no error cancellation. It should be noted that even with λ = 0.01 or λ = 0.30, the use of the same protein and water coordinates for both ligands introduces a systematic error for the interactions with the transformed fluorine atom of ligand 4. This explains why the individual corrections to ΔG_{bound} and ΔG_{free} are so different from those obtained with Eq. 6. The systematic error seems to cancel rather well between ΔG_{bound} and ΔG_{free}, but this could be somewhat fortuitous. In conclusion, the approximate endpoint methods can provide a better understanding of the problem, but are not quantitatively reliable.

We now investigate why the QM free energies calculated by BAR show much smaller statistical errors than the endpoint corrections obtained by exponential averaging (cf. Table 2), despite employing almost the same data. Loosely speaking, the two methods follow different paths between the same endpoints (although one should keep in mind that the BAR method is not rigorous). Although the transitions along the BAR path are also suffering from limited sampling, they are not dominated by a single snapshot. If one would calculate these free energies by exponential averaging, the weight of the largest term in the exponential sum would be 0.98 and 0.71 for the most difficult forward (λ = 0.30 → λ = 0.99) and backward (λ = 0.99 → λ = 0.30) transitions, respectively, and the use of BAR further reduces the statistical problems as it combines data from both ensembles. As shown in Figure S4 in the Supporting information, there is significant overlap between the energy distributions of the three ensembles, indicating that similar geometries are sampled in the three cases, except for some repulsive structures in the λ = 0.99 ensemble which will not contribute to the free energy in any case. In conclusion, the results indicate that the path taken in the BAR method (employing only energies at the QM level) avoids the explicit transitions between the QM and MM potential-energy surfaces, which seem to differ significantly for this system.

Basis-set corrections

The basis set used in the large set of QM calculations (def2-SV(P)) is far from saturated. To estimate the missing effect, we performed QM calculations with a larger basis set (def2-TZVP) for a subset of the snapshots, and calculated corrections for the basis-set superposition error (BSSE) for both basis sets using the counterpoise procedure [90]. The resulting corrections are shown in Table 3. As expected, there is a substantial BSSE (20–44 kJ/mol for relative interaction energies) when using the smaller basis set, but this is reduced to 1–5 kJ/mol with the larger basis set. On the other hand, there is a large effect (13–27 kJ/mol) from the change of the basis set itself.

For both the bound and free states, the geometries used to calculate the basis-set corrections were taken from each of the three ensembles of snapshots (λ=0.01, 0.30, and 0.99). Unfortunately, the

number of snapshots (11 per ensemble) was too small to reliably calculate the BAR results with the larger basis set, so the basis-set corrections in Table 3 are plain averages. It is therefore satisfying to note that regardless of how the geometries were chosen, a positive basis-set correction to the total relative binding free energy was obtained. The most natural approach, analogous to Eq. 6, would be to use different sets of geometries for **3** and **4** (λ=0.01 and λ=0.99, respectively); this approach is denoted *mixed* in Table 3. However, the use of different geometries ruins the error cancellation between the two ligands and leads to a standard error of 15 kJ/mol, which renders the computed average rather useless. Therefore, the only viable approach is to compute the basis-set correction with the same geometry for the two ligands, analogous to Eq. 7. Using the intermediate (λ=0.30) geometries for this purpose, to avoid bias to any of the end points, we estimate the correction to 10±2 kJ/mol.

Effect of the surroundings

The restriction of the QM calculations to atoms within 6 Å of the ligand could potentially be a severe approximation. The effect of the remaining atoms, i.e. the surrounding protein (for the bound state) and water molecules, was estimated using the polarizable multipole interaction (PMI) method. This method includes the direct interactions between the ligand and the surrounding system, but also the indirect (many-body) effects through a description of the whole system as a collection of polarizable multipoles. The results are given in Table 4. As in the case of the QM effect, differences based on free-energy evaluations (BAR or NBB) as well as plain averages of interaction energies are given. For comparison, the corresponding results using the simple pairwise-additive Amber potential are also shown.

If one disregards the NBB estimates, which again suffer from a large statistical uncertainty, all the results indicate that the surroundings stabilise ligand **4** in the bound state, whereas the surroundings stabilise ligand **3** in the free state. For the bound state, the magnitude of this effect varies with method; it is 4–5 kJ/mol greater when evaluated by the PMI method than with Amber, and it is 5 kJ/mol greater when evaluated by plain averages than by BAR, regardless of the potential. For the free state, the effect is 1–3 kJ/mol and it is largely independent of the method used.

From a computational point of view, it is interesting to know how many water molecules that need to be included in the PMI calculations. In Figure 6, the calculated ΔG_{bound} and ΔG_{free} at the QM/PMI level (Eq. 9) are plotted as a function of the distance R from the ligand, within which all water molecules are included (note that water molecules within 6 Å of the ligand are already included in the QM system). Compared to the effect of the surrounding protein (cf. Table 4), the additional effect of far-lying water molecules is rather weak; the results are converged (within 1 kJ/mol) at a distance of 8–20 Å and even neglecting all water molecules outside the QM system causes an error of only 2 kJ/mol. The total contribution from the surroundings to the relative binding free energy at the BAR level is 9 kJ/mol, favouring the binding of ligand **4**.

Accuracy of the final result

The total QM/PMI relative binding free energy evaluated with BAR is –4±2 kJ/mol (the difference of the ΔG_{bound} and ΔG_{free} results at R = 28 Å in Figure 6), as summarised in Table 5. When we add the corrections due to the incomplete basis set and the restriction to interaction energies, we arrive at a final relative binding free energy of 3±3 kJ/mol, i.e. practically identical to our original estimate at the MM level (3.6±0.5 kJ/mol). Thus, there is still a ~7 kJ/mol discrepancy between the

calculated and the experimental value (10 kJ/mol [64]).

There are at least four sources of error in the calculations that could contribute to this discrepancy. First, we found a large dependence on the basis set used in the QM calculations. Unfortunately, this basis-set dependence did not cancel between the bound and free simulations, but introduced a systematic shift in the relative binding free energy. Although we performed a correction to the triple-zeta level, where the BSSE is rather small, it is possible that further enhancement of the basis-set quality would cause an additional systematic shift.

Second, there is an inherent error in the DFT-D method. Grimme and coworkers have reported average errors in interaction energies calculated by DFT-D of 4–8 kJ/mol or <5% of the interaction energies [101,102]. We have investigated this issue for another system by performing local CCSD(T) calculations extrapolated to the complete basis-set limit [18]. However, the computational cost of such high-level QM methods limits their applicability to very few snapshots, and thus makes the statistical evaluation even more problematic than for the basis-set convergence. Moreover, it is difficult to compare the DFT-D errors between different systems, because error cancellation plays a large role and hence the error depends on how similar the two ligands are.

Third, the QM free-energy evaluation using BAR is based on an approximate theory, in which the QM interaction energies are computed on snapshots obtained by MD using a standard MM force field. From the more rigorous NBB calculations, we infer that the BAR results are reasonably accurate, but owing to the poor precision of the NBB results, we cannot exclude the possibility that the BAR estimates are biased by several kJ/mol in any direction. Increasing the number of snapshots and especially the number of λ values would improve the NBB precision and thus give an answer to this question.

Finally, it is also possible that the discrepancy between the experimental and computational results is not caused by shortcomings of the MM force field, but rather from an incomplete sampling of the phase space, e.g. caused by conformational changes or differences in binding mode between the two ligands [2]. For galectin-3, two different conformations of Arg-144 have been observed in both crystal structures and MD simulations [103,104], above or below the benzene group of the ligand. Such problems can only be solved by much longer simulations or methods that locally enhance the sampling [105].

Conclusions

We have computed the relative free energy of two synthetic disaccharide ligands binding to galectin-3 using a combination of molecular dynamics with a standard MM force field followed by 3600 single-point QM evaluations, in which a central system (all atoms within 6 Å; i.e. ~527 or 748 atoms) was treated by dispersion-corrected DFT, and the rest of the system was treated by an accurate polarizable multipole description. The purpose of the study was to investigate how to compute a statistically converged binding free energy with a reasonably high-level QM/MM method and a proper sampling. This question has not been thoroughly discussed before, because previous studies have either used a simpler QM protocol (e.g. treating only the ligand by QM) or assumed a single (minimum) structure for the protein–ligand complex.

We found that the MM and QM potential-energy surfaces differ to such an extent that endpoint corrections based on exponential averaging do not work, because they are completely dominated by a single data point. Instead, we investigated two approaches, in which QM calculations for three λ

values were performed. Unfortunately, the most rigorous of these approaches, NBB, gave too large statistical errors and could only be used to establish that the other approach, the direct use of the BAR method with QM energies, seems to work reasonably well.

The effect on $\Delta\Delta G$ of switching from MM to QM was found to be 5 kJ/mol. However, the effect of switching basis set from double- to triple-zeta quality is even larger, 10 kJ/mol. The effect from protein residues and water molecules outside the central system was found to be 9 kJ/mol, whereas the effect of using interaction energies instead of total energies in the alchemical transformation is 3 kJ/mol. Thus, none of these effects are negligible. However, the various effects incidentally cancel out, so that our best QM estimate of $\Delta\Delta G$ is almost identical to the MM estimate and still deviating from the experimental value by 7 kJ/mol.

The statistical uncertainty of the final result is 3 kJ/mol and is dominated by the correction to the larger basis set. Thus, the uncertainty can probably be reduced to less than 2 kJ/mol by performing all the QM calculations with the larger basis set, albeit at a significantly higher computational cost. However, we cannot exclude the possibility that an even larger basis set is needed. It should also be noted that the QM-corrected BAR method is not a rigorous method; it relies on the assumption that the QM and MM potential-energy surfaces have similar shape. The validity of this assumption can be expected to vary from system to system. Moreover, it is possible that insufficient sampling of the conformational space may limit the accuracy of the calculated affinities.

Nevertheless, the results demonstrate that AP can be combined with a QM reweighting scheme that gives a statistical precision of 2–3 kJ/mol, while using ~750 atoms in the QM system coupled with an accurate description of the surroundings. This is a promising step towards the computation of statistically converged binding free energies without empirically fitted parameters. Such computation would finally allow an unbiased comparison between experimental binding free energies and the corresponding theoretical predictions.

Acknowledgements

This investigation has been supported by grants from the Swedish research council (agreement C0020401 and project 2010-5025) and the Research School in Pharmaceutical Science (FLÄK). The computations were performed on computer resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lunarc at Lund University and HPC2N at Umeå University.

References

[1] M. K. Gilson, H.-X. Zhou, *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42
[2] D. L. Mobley, P. V. Klimovich, *J. Chem. Phys.* **2012**, *137*, 230901
[3] R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420-1426
[4] J. G. Kirkwood, *J. Chem. Phys.* **1935**, *3,* 300–313
[5] C. H. Bennett, *J. Comput. Phys.* **1976** *22,* 245–268
[6] M. R. Shirts, V. S. Pande, *J. Chem. Phys.* **2005**, *122*, 144107
[7] K. M. J. Merz, *J. Chem. Theory Comput.* **2010**, *6*, 1018-1027
[8] D. A. Pearlman, P. S. Charifson, *J. Med. Chem.* **2001**, *44*, 3417
[9] S. Genheden, I. Nilsson, U. Ryde, *J. Chem. Inf. Model.* **2011**, *51*, 947-958
[10]S. Genheden, P. Mikulskis, U. Ryde, *J. Chem. Inf. Model.,* **2014**, *54*, 2794-2806
[11]D. A. Pearlman, *J. Med. Chem.* **2005**, *48*, 7796
[12]G. J. Rocklin, D. L. Mobley, K. A. Dill, *J. Chem. Theory Comput.* **2013**, *9*, 3072-3083
[13]H. S. Muddana, M. K. Gilson, *J. Comput. Aided Mol. Des.* **2012**, *26*, 517-525
[14]D. Jiao, P. A. Golubkov, T. A. Darden, P. Y. Ren, *Prot. Nat. Acad. Sci. USA* **2008**, *105*, 6290
[15]J. J. Zhang, W. Yang, J.-P. Piquemal, P. Y. Ren, *J. Chem. Theory Comput.* **2012**, *8*, 1314-1324
[16]P. Söderhjelm, S. Genheden, U. Ryde, in *Protein-Ligand Interactions*; Gohlke, H, Ed.; Wiley-VCH: Weinheim, **2012**; Chapter 7, p 121-144
[17]Y. Mochizuki, K. Yamashita, T. Nakano, Y. Okiyama, K. Fukuzawa, N. Taguchi, S. Tanaka, *Theor. Chem. Acc.* **2011**, *130*, 515-530
[18]M. Andrejic, U. Ryde, R. Mata, P. Söderhjelm, *ChemPhysChem* **2014**, 15, 3270-3281
[19]P. Söderhjelm, F. Aquilante, U. Ryde, *J. Phys. Chem. B* **2009**, *113*, 11085
[20]T. Sawada, D. G. Fedorov, K. Kitaura, *J. Phys. Chem. B* **2010**, *114*, 15700-15705
[21]A. Morreale, F. Maseras, I. Iriepa, E. Gálvez, *J. Mol. Graph. Model.* **2002**, *21*, 111-118.
[22]L. C. Menikarachchi, J. A. Gascon, *Current Topics in Med. Chem.* **2010**, *10*, 46-54
[23]J. Antony, S. Grimme, *J. Comput. Chem.* **2012**, *33*, 1730-1739.
[24]J. Fanfrlík, J. Brynda, J. Rezác, P. Hobza, M. Liepsík, *J. Phys. Chem. B* **2008**, *112*, 15094-05102
[25]C. A. Morgado, I. H. Hillier, N. A. Burton, J. J. McDouall, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2706-2714
[26]R. Lonsdale, J. N. Harvey, A. J. Mulholland, *J. Chem. Theory Comput.* **2012**, *8*. 4637-4645
[27]M. R. Reddy, M. D. Erion, *J. Am. Chem. Soc.* **2007**, *129*, 9296-9297.
[28]R. S. Rathore, R. N. Reddy, A. K. Kondapi, P. Reddanna, M. R. Reddy, *Theor. Chem. Acc.* **2012**, *131*, 1096
[29]F. Gräter, S. M. Schwarzl, A. Dejaegere, S. Fischer, J. C. Smith, *J. Phys. Chem. B* **2005,** *109*, 10474-10483
[30]A. Khandelwal, V. Lukacova, D. Comez, D. M. Kroll, S. Raha, S. Balaz, *J. Med. Chem.* **2007**, *48*, 5437-5447.
[31]M. Retegan, A. Milet, H. Jamet, *J. Chem. Inf. Model.* **2009**, 49, 963-971
[32]P. Söderhjelm, J. Kongsted, U. Ryde, *J. Chem. Theory Comput.* **2010**, *6*, 1726 -1737
[33]S. Genheden, T. Luchko, S. Gusarov, A. Kovalenko, U. Ryde, *J. Phys. Chem. B*, **2010**, *114*, 8505-8516
[34]S. Genheden, P. Mikulskis, L. Hu, J. Kongsted, P. Söderhjelm, U. Ryde, *J. Am. Chem. Soc.*, **2011**, 133, 13081-13092
[35]S. Genheden, U. Ryde, *Expert Opion Drug Design*, submitted, **2015**.
[36]V. Luzhkov, A. Warshel, *J. Comput. Chem.* **1992**, *13*, 199-213
[37]R. P. Muller, A. Warshel, *J. Phys. Chem.* **1995**, *99*, 17516-17524
[38]R. Iftimie, D. Salahub, D. Wei, J. Schofield, *J. Chem. Phys.* **2000**, *113*, 4852
[39]P. Bandyopadhyay, *J. Chem. Phys.* **2005**, *122*, 091102
[40]T. Wesolowski, A. Warshel, *J. Phys. Chem.* **1994**, *98*, 5183-5187
[41]M. H. Olsson, G. Hong, A. Warshel, *J. Am. Chem. Soc.* **2003**, *125*, 5025-5039
[42]R. H. Wood, E. M. Yezdimer, S. Sakane, J. A. Barriocanal, D. J. Doren, *J. Chem. Phys.* **1999**, *110*, 1329
[43]T. H. Rod, U. Ryde, *Phys. Rev. Lett.* **2005**, *94*, 138302
[44]N. V. Plotnikov, S. C. L. Kamerlin, A. Warshel, *J. Phys. Chem. B* **2011**, *115*, 7950-7962
[45]F. R. Beierlein, J. Michel, J. W. Essex, *J. Phys. Chem. B* **2011**, *115*, 4911-4926
[46]S. J. Fox, C. Pittock, C. S. Tautermann, T. Fox, C. Christ, N. O. J. Malcolm, J. W. Essex, C.-K. Skylaris, *J. Phys. Chem. B* **2013**, *117*, 9478-9485
[47]M. Strajbl, G. Hong, A. Warshel, *J. Phys. Chem. B* **2002**, *106*, 13333-13343
[48]N. V. Plotnikov, A. Warshel, *J. Phys. Chem. B* **2012**, *116*, 10342-10356
[49]J. Heimdal, U. Ryde, *Phys. Chem. Chem. Phys.* **2012**, *14*, 12592-12604
[50]K. E. Shaw, C. J. Woods, A. J. Mulholland, *Phys. Chem. Lett.* **2010**, *1*, 219-223
[51]C. J. Woods, F. R. Manby, A. J. Mulholland, *J. Chem. Phys.* **2008**, *128*, 014109
[52]C. J. Woods, K. E. Shaw, A. J. Mulholland, *J. Phys. Chem. B* **2015***, 119*, 997-1001
[53]Y. Zhang, H. Liu, W. Yang, *J. Chem. Phys.* **2000**, *112*, 3483
[54]J. Kästner, H. M. Senn, S. Thiel, N. Otte, W. Thiel, *J. Chem. Theory Comput.* **2006**, *2*, 452-461
[55]G. König, S. Boresch, *J Comput Chem* **2011**, *32*, 1082-1090

[56]G. König, P. Hudson, S. Boresch, H. L. Woodcock, *J. Chem. Theory Comput.* **2014**, 10, 1406-1419
[57]G. König, F. C. Pickard 4th, Y. Mei, B. R. Brooks, *J. Comput. Aided Mol. Des.* **2014**, *28*, 245-257
[58]P. Mikulskis, D. Cioloboc, M. Andrejic, S. Khare, J. Brorsson, S. Genheden, R. A. Mata, P. Söderhjelm, U. Ryde, *J. Comput. Aided Mol. Des.* **2014**, 28, 375-400
[59]H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198-1229
[60]P. Söderhjelm, U. Ryde, *J. Phys. Chem. A* **2009**, *113*, 617-627
[61]J. H. Jensen, H. Li, A. D. Robertson, P. A. Molina, *J Phys Chem A* **2005**, *109*, 6634-6643
[62]B. de Courcy, J.-P. Piquemal, C. Garbay, N. Gresh, *J. Am. Chem. Soc.* **2010**, *132*, 3312-3320
[63]O. Engkvist, P.-O. Åstrand, G. Karlström, *Chem. Rev.* **2000**, *100*, 4087
[64]P. Sörme, P. Arnoux, B. Kahl-Knutsson, H. Leffler, J. M. Rini, U. J. Nilsson, *J. Am. Chem. Soc.* **2005**, *127*, 1737-1743
[65]Steinbrecher, T.; Mobley, D. L.; Case, D. A. *J. Chem. Phys.* **2007**, *127*, 214108.
[66]Steinbrecher, T.; Joung, I.; Case, D.A. *J. Comp. Chem.*, **2011**, *32*, 3253–3263.
[67]M. R. Shirts, J. D. Chodera, *J. Chem. Phys.* **2008**, *129*, 124105
[68]S. Genheden, U. Ryde, *J. Comput. Chem.* **2010**, *31*, 837– 846
[69]J. Uranga, P. Mikulskis, S. Genheden, U. Ryde, *Comput. Theor. Chem.* **2012***, 1000*, 75-84.
[70]V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins: Struct., Funct. Bioinform.* **2006**, *65*, 71
[71]J. M. Wang, R. M. Wolf, K. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157– 1174
[72]C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269– 10280
[73]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
[74]L. Hu, P. Söderhjelm, U. Ryde, *J. Chem. Theory Comput.* **2013**, *9*, 640-649
[75]D. A. Case, T. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R. Woods, *J. Comput. Chem*. **2005**, *26*, 1668-1688.
[76]J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327– 341
[77]M. R. Shirts, D. L. Mobley, J. D. Chodera, V. S. Pande, *J. Phys. Chem. B*, **2007**, *111*, 13052–13063
[78]T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089– 10092
[79]X. Wu, B. R. Brooks, *Phys. Lett.* **2003**, *381*, 512– 518
[80]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690
[81]R. Ahlrichs, M. Bär, M. Häser, H. Horn, K. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165
[82]O. Treutler, R. Ahlrichs, *J. Chem. Phys.* **1995**, *102*, 346
[83]A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098-3100.
[84]C. T. Lee, W. T. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785
[85]A. Schäfer, C. Huber, R. Ahlrichs, *J. Chem. Phys.* **1994**, *100*, 5829
[86]S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J Chem Phys* **2010**, *132*, 154104
[87]J. Antony, S. Grimme, *Phys. Chem. Chem. Phys.*, **2006**, *8*, 5287–5293
[88]K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *240*, 283-290
[89]K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs, *Theor. Chem. Acc.* **1997,** *97*, 119-126
[90]S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553-566
[91]F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305
[92]http://www.thch.uni-bonn.de/tc/index.php?section=downloads&subsection=getd3
[93]L. Gagliardi, R. Lindh, G. Karlström, *J. Chem. Phys.* **2004**, *121*, 4494
[94]MOLCAS 7, University of Lund, Sweden
[95]F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-Å. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitonak, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov, R. Lindh, *J. Comp. Chem.* **2010**, *31*, 224
[96]D. W. Zhang, J. Z. H. Zhang, *J. Chem. Phys.* **2003**, *119*, 3599
[97]P. Politzer, K. E. Riley, F. A. Bulat, J. S. Murray, *Comput. Theor. Chem.* **2012**, *998*, 2-8.
[98]W. L. Jorgensen, P. Schyman, *J. Chem. Theory Comput.* **2012**, *8*, 3895-3901.
[99]W. Liu, S. Sakane, R. H. Wood, D. J. Doren, *J. Phys. Chem. A* **2002**, *106*, 1409
[100]C. Cave-Ayland, C.-K. Skylaris, J. W. Essex, *J. Phys. Chem. B* **2015**, *119*, 1017−1025
[101]J. Antony, S. Grimme, D. G. Liakos, F. Neese, *J. Phys. Chem. A* **2011**, *115*, 11210-11220.
[102]S. Grimme, *Chem. Eur. J.* **2012**, *18,* 9955
[103]C. Diehl, O. Engstrom, T. Delaine, M. Hakansson, S. Genheden, K. Modig, H. Leffler, U. Ryde, U. J. Nilsson, M. Akke, *J. Am. Chem. Soc.* **2010**, *132*, 14577.
[104]F. Godschalk, S. Genheden, P. Söderhjelm, U. Ryde, *Phys. Chem. Chem. Phys.*, **2013**, *15*, 7731-7739
[105]S. A. Adock, J. A. McCammon, *Chem. Rev.* **2006**, *106*, 1589-1615.

Table 1. Free energies (kJ/mol) of transforming **3** to **4** in the protein (ΔG_{bound}), in solution (ΔG_{free}), and the difference ($\Delta\Delta G$), i.e. the total relative binding free energy, calculated by BAR at the MM level. Precise results using eleven λ values and approximate results using three λ values are given, as well as results using total energies, interaction energies using all atoms, and interaction energies using system S_1, respectively (only $\Delta\Delta G$ is comparable between total and interaction energies).

	ΔG_{bound}	ΔG_{free}	$\Delta\Delta G$
Total (11 λ)	-28.7±0.3	-32.3±0.4	3.6±0.5
Total (3 λ)	-28.3±0.9	-33.7±1.0	5.4±1.3
Interaction (11 λ)	18.3±0.2	11.6±0.2	6.7±0.2
Interaction (3 λ)	19.6±0.3	13.2±0.3	6.4±0.4
S_1 interaction (11 λ)	20.1±0.2	9.4±0.2	10.7±0.2
S_1 interaction (3 λ)	21.6±0.3	11.2±0.3	10.4±0.5

Table 2. Free energies (kJ/mol) calculated by BAR and NBB at the MM and QM levels (with the same columns as in Table 1) for system S_1, but using the MM energy of the full system as the reference level in the NBB calculations (i.e. as V_{MM} for calculating V_{bias} in Eq. 8). The difference between the QM and MM result with the same methods are also given, together with three ways of estimating this difference from only the simulations at the endpoints: the rigorous exponential averaging (Eq. 5), the approximation to use plain averages instead of exponential averages (Eq. 6), and the same approximation but using only the snapshots from the λ=0.01 or λ=0.30 simulation (Eq. 7). QM results are obtained at the BLYP-D3/def2-SV(P) level without any counterpoise corrections.

	ΔG_{bound}	ΔG_{free}	ΔΔG
MM			
BAR	21.6±0.3	11.2±0.3	10.4±0.5
NBB	24.9±1.9	11.2±4.6	13.7±5.0
QM			
BAR	34.6±0.8	29.3±1.1	5.2±1.4
NBB	36.0±4.7	21.8±11.6	14.2±12.5
NBB indirect	68.1±15.3	45.4±10.9	22.7±18.8
Difference			
BAR	13.0±0.8	18.2±1.2	-5.1±1.4
NBB	11.1±5.1	10.6±12.5	0.5±13.5
endpoint, exp	53.5±17.6	34.7±14.1	18.8±22.6
endpoint, plain	12.6±4.8	24.2±6.6	-11.6±8.1
endpoint, plain, λ=0.01	21.3±0.7	33.5±0.8	-12.2±1.1
endpoint, plain, λ=0.30	22.6±0.8	32.8±1.3	-10.2±1.5

Table 3. BSSE corrections (difference between counterpoise-corrected and non-corrected interaction energies) at the BLYP/def2-SV(P) and BLYP/def2-TZVP levels, as well as the basis-set correction, i.e. the difference in the non-corrected interaction energy between BLYP/def2-TZVP and BLYP/def2-SV(P), and the total correction, which is the sum of the basis-set correction and the BSSE correction at the BLYP/def2-TZVP level. All values are averages over 21 (BSSE corr. with SVP) or 11 snapshots (others) with standard error σ and all values are differences between the V_1 and V_0 corrections. The snapshots are either taken from one simulation (λ=0.01, 0.30, or 0.99) or from different simulations (λ=0.01 for V_0 and λ=0.99 for V_1; denoted *mixed*). The results for the bound and free states are reported separately, and the difference between them, i.e. the net correction contributing to ΔΔG, is also given. All energies are in kJ/mol.

	ΔG_{bound}				ΔG_{free}				$\Delta\Delta G$			
	λ=0.01	λ=0.30	λ=0.99	mixed	λ=0.01	λ=0.30	λ=0.99	mixed	λ=0.01	λ=0.30	λ=0.99	mixed
BSSE corr. (SVP)	-20.2	-21.5	-24.4	-12.6	-27.8	-31.4	-43.8	-24.3	7.6	9.9	19.5	11.7
σ	1.1	1.0	2.7	7.4	1.2	1.4	1.8	8.4	1.6	1.7	3.2	11.2
BSSE corr. (TZVP)	-2.0	-2.0	-1.2	-1.1	-4.2	-4.4	-5.3	-5.7	2.2	2.4	4.0	4.6
σ	0.2	0.2	1.3	1.5	0.2	0.3	0.3	1.2	0.3	0.3	1.3	1.9
Basis-set corr.	-20.0	-19.5	-13.1	-15.9	-26.0	-27.3	-20.2	-32.2	6.0	7.8	7.1	16.3
σ	2.1	1.1	1.9	11.6	2.2	1.4	5.8	7.9	3.0	1.8	6.1	14.1
Total correction	-22.0	-21.5	-14.3	-17.0	-30.2	-31.7	-25.5	-37.9	8.3	10.1	11.2	20.9
σ	2.2	1.3	2.6	12.8	2.3	1.6	5.6	8.7	3.2	2.1	6.2	15.5

Table 4. Effect of the surroundings (difference between the free energies evaluated for S_2 and S_1) on ΔG_{bound} and ΔG_{free}, calculated using PMI or Amber for the same free-energy methods as in Table 2 (the endpoint exponential-averaging result was omitted owing to the poor precision). The water-inclusion radii for defining S_2 were R = 6 Å for ΔG_{bound} and R = 22 Å (all water) for ΔG_{free}. All energies are in kJ/mol.

	ΔG_{bound}		ΔG_{free}	
	PMI	Amber	PMI	Amber
BAR	-6.9±1.1	-2.8±1.1	2.3±1.6	1.5±1.6
NBB	-11.1±6.0	0.3±6.8	12.5±14.7	7.4±14.8
endpoint, plain	-12.2±1.1	-7.5±1.0	1.1±3.7	3.3±2.7
endpoint, plain, λ=0.01	-8.3±0.3	-4.8±0.3	2.8±0.6	3.1±0.6
endpoint, plain, λ=0.30	-6.4±0.3	-2.2±0.3	2.0±0.6	2.0±0.5

Table 5. Summary of contributions (kJ/mol) to the estimated $\Delta\Delta G$ at the QM/PMI level. For comparison, the Amber result and the experimental value is also given.

	ΔΔG
Total QM/PMI [a]	-4.1±1.5
QM for S_1	5.2
PMI effect [b]	-9.3
Basis-set corrections	10.1±2.1
Interaction correction[c]	-3.1±0.5
Final QM/PMI result	**2.9±2.6**
Amber result	3.6±0.5
Experiment [64]	**10.3**

[a] Using the largest considered S_2, i.e. R = 28 Å for the bound state and R = 22 Å (all water) for the free state (cf. Figure 6).
[b] Defined as the difference between the QM/PMI and QM for S_1 results.
[c] Difference at the MM level between ΔΔG calculated using total energies and interaction energies, respectively (cf. Table 1).

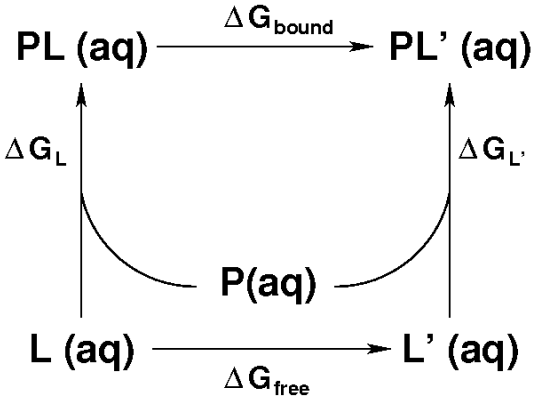


Figure 1. The thermodynamic cycle used to estimate the relative binding affinity $\Delta G_{L'} - \Delta G_{L}$.

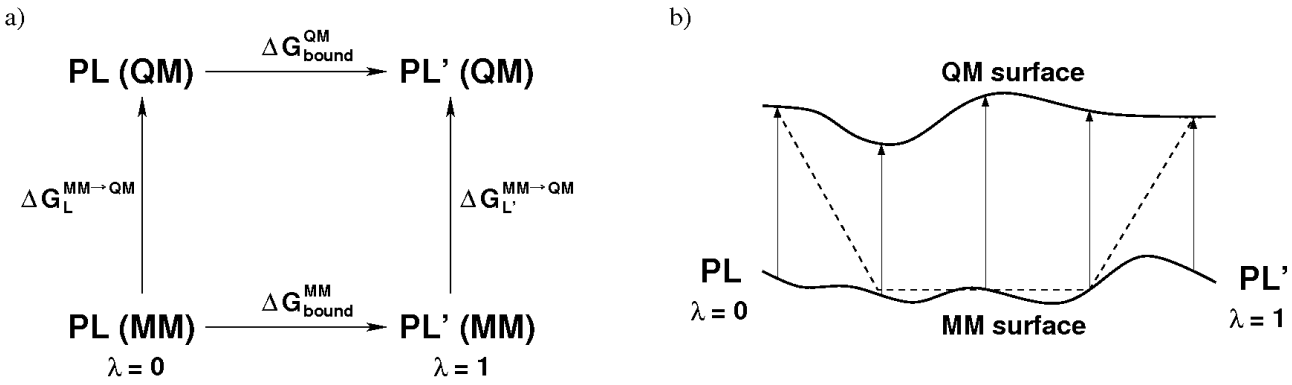


Figure 2. The two types of methods used for estimating ΔG_{bound} at the QM level. a) The thermodynamic cycle used in the endpoint method. b) Schematic illustration of the NBB method; the arrows represent the use of the ensembles generated for various λ values at the MM surface to perform the free-energy evaluation at the QM surface. The dashed line symbolizes the indirect version of the method [56,57], in which the QM energy function is only used for the first and last λ values, whereas the MM energy function is used for the intermediate λ values.

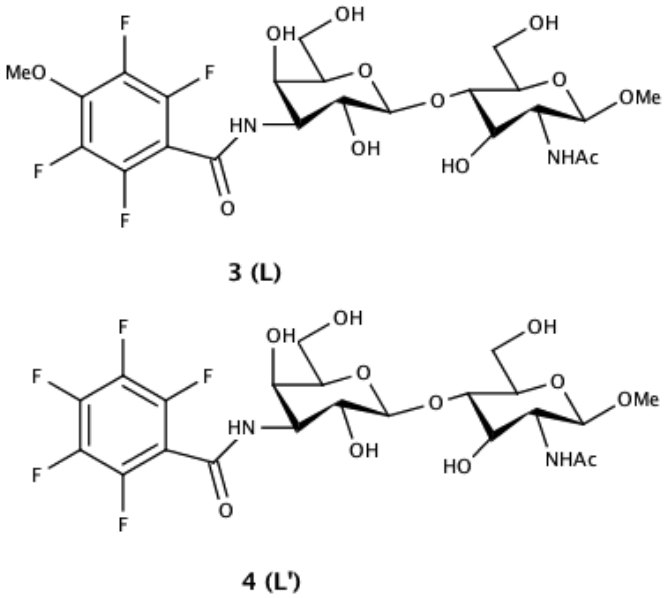


Figure 3. The two gal3 ligands considered in the calculations.

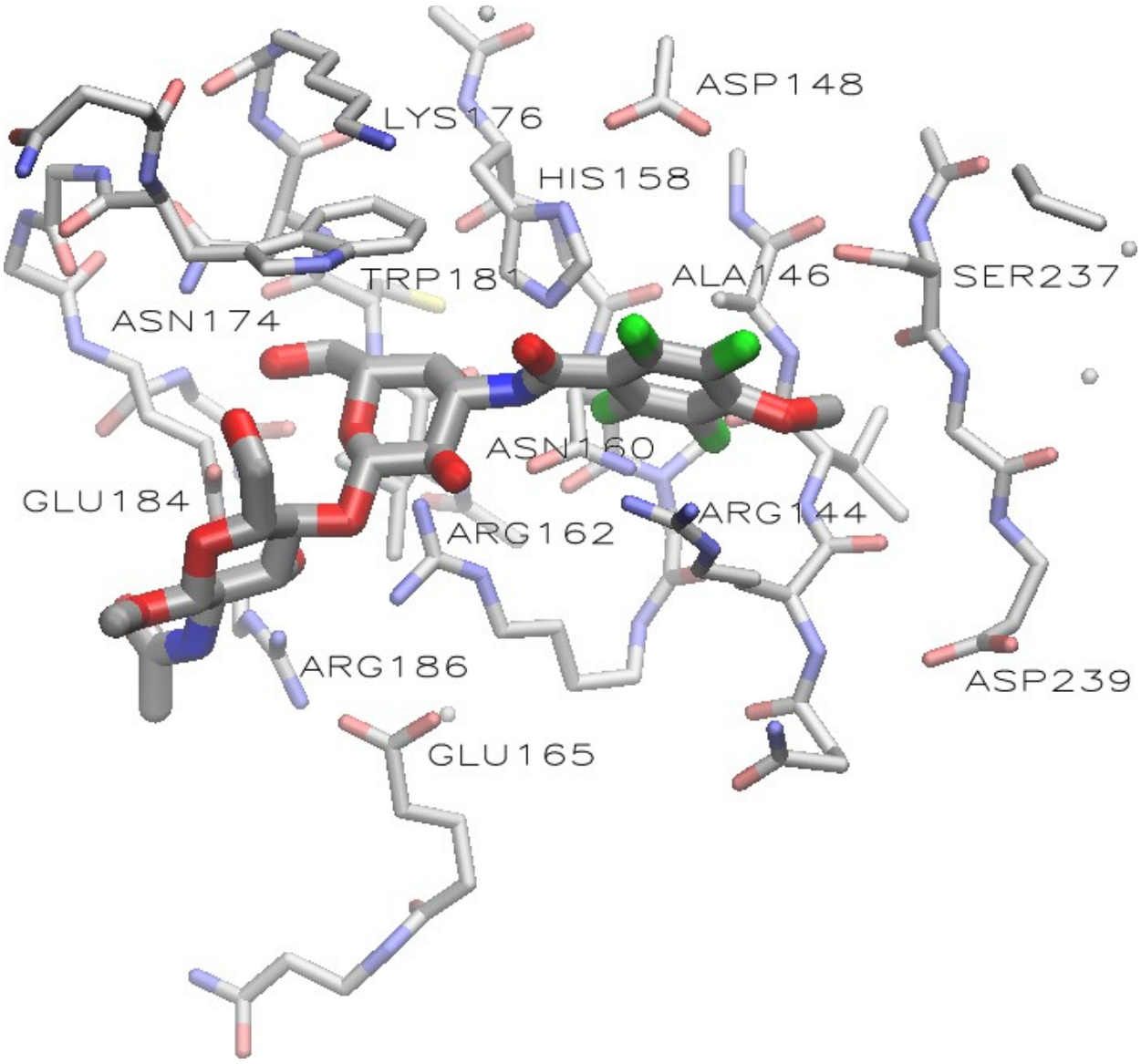


Figure 4. System S_1, i.e. atoms included in the QM calculations of the bound state. The ligand (**3** in this case) is drawn with thicker bonds and important protein residues (either charged or close-lying) are labelled. 97 water molecules were also included in the calculations, but they are omitted in the figure because their positions vary significantly between the snapshots. Hydrogen atoms are also omitted for clarity. The small balls represent single carbons, i.e. methane molecules.

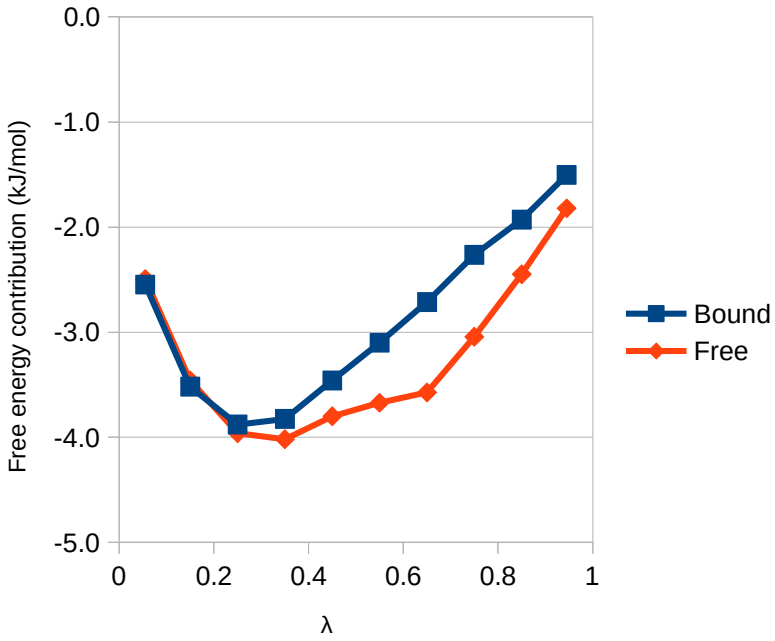


Figure 5. Contributions to ΔG_{bound} and ΔG_{free} from each of the (λ_i, λ_{i+1}) intervals, represented by the midpoint of the interval.

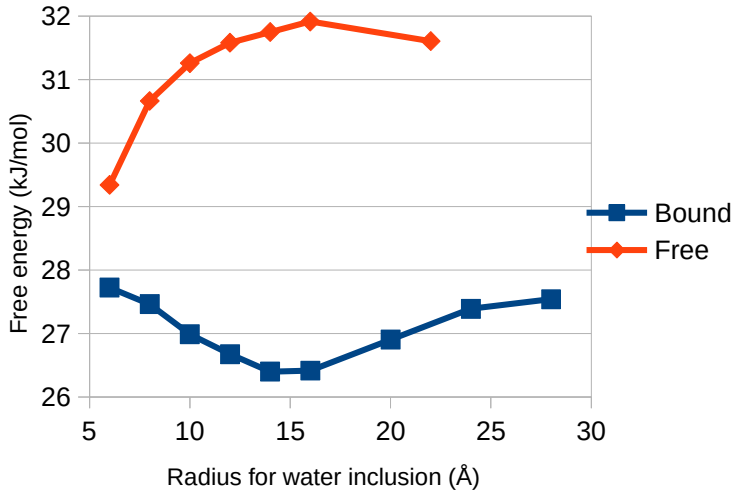
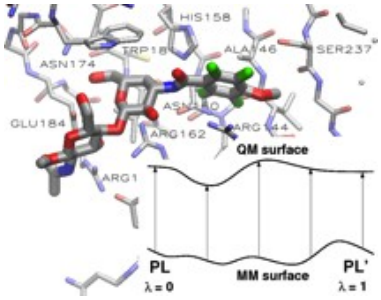


Figure 6. BAR results for ΔG_{bound} and ΔG_{free} evaluated by QM/PMI (Eq. 9) as a function of the radius R within which water molecules are included in the PMI treatment. For ΔG_{free}, all water molecules are included at $R = 22$ Å.

Graphical Abstract to

Samuel Genheden, Ulf Ryde, and Pär Söderhjelm: *Binding affinities by alchemical perturbation using QM/MM with a large QM system and polarizable MM model*



The binding of two ligands to galectin-3 is used as a test case for exploring methods that combine alchemical perturbation with quantum-mechanical energy evaluations.

Supporting information

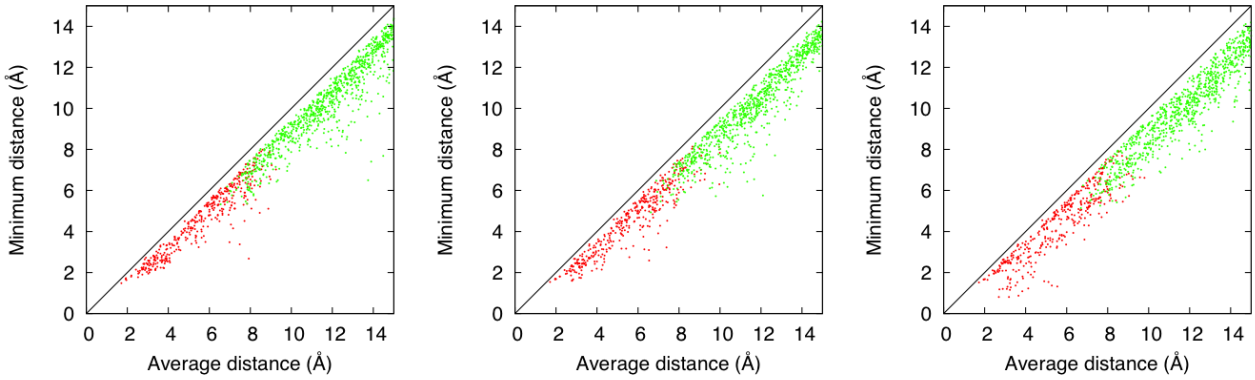


Figure S1. Minimum distance from each protein atom to the ligand plotted against the average distance over 100 snapshots from the λ=0.01 (left), 0.30 (centre), and 0.99 (right) simulations. The red points denote atoms in S_1 and the green points denote atoms outside S_1. The plots have been truncated at 15 Å. The closest atom outside S_1 (always Pro-5:HA) has a minimum distance of 5.3, 5.0, and 5.1 Å in the three simulations, and an average distance of 6.6, 6.6, and 6.7 Å.

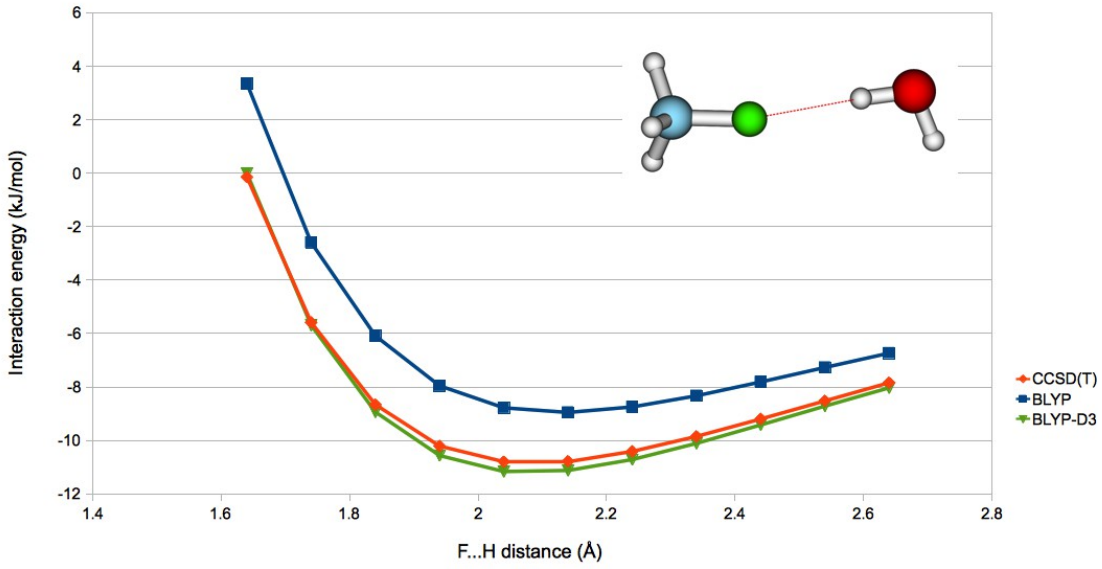


Figure S2. Benchmark of the BLYP functional against high-level QM calculations for the fluoromethane–water dimer optimized at the MP2/6-311++G(2d,2p) level (see inset) and a set of geometries obtained by decreasing and increasing the F...H distance in steps of 0.1 Å while keeping all other internal coordinates fixed at the optimized geometry. For the reference CCSD(T) calculations, the aug-cc-pVTZ basis set was used. For the BLYP calculations, the def2-TZVP basis set was used, i.e. the target level for the QM free energies calculated in this study. Counterpoise corrections were used in both sets of calculations. The green curve shows the BLYP results after addition of the DFT-D3 dispersion correction (with default parameters) also employed throughout this study. The maximum deviation from the CCSD(T) result in this interval is 0.4 kJ/mol, which is probably somewhat fortuitous, but indicates that the employed QM methodology is adequate for the type of interactions occurring in the studied protein-ligand system.

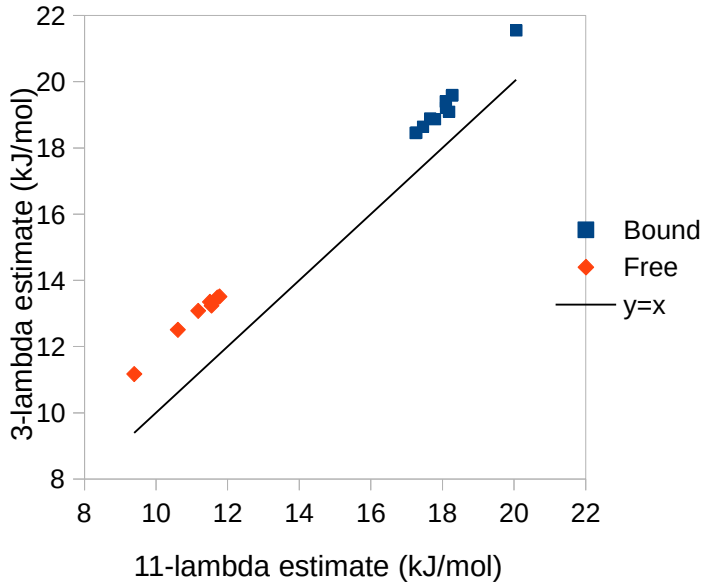


Figure S3. Correlation between the precise (11-lambda) and approximate (3-lambda) estimates of ΔG_{bound} and ΔG_{free} from BAR calculations using different system sizes. The 1–2 kJ/mol error appears largely systematic; thus an even smaller error is obtained when taking the difference between ΔG_{bound} and ΔG_{free}, independently of the chosen system size.

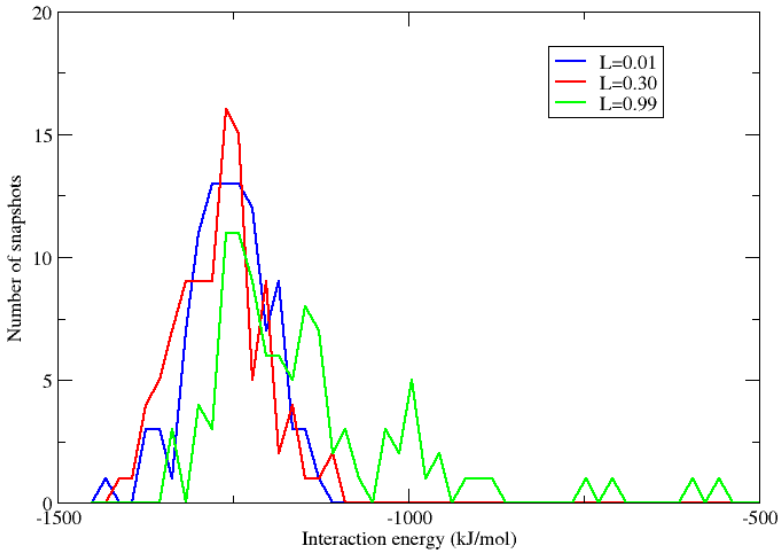


Figure S4. Distribution of $V^{QM}_{\lambda=0.30}$ (see definition below) for system S_1 in the three ensembles (λ=0.01, 0.30, and 0.99, respectively).

Soft-core potentials

For all interactions with the atoms in the disappearing –OCH_3 group in **3**, the interaction-energy term is modified into [65,66]

$$V_0(\text{disappearing}) = 4\varepsilon\left[\frac{1}{\left(\alpha\lambda+(r_{ij}/\sigma)^6\right)^2} - \frac{1}{\alpha\lambda+(r_{ij}/\sigma)^6}\right] + \frac{q_i q_j}{4\pi\epsilon_0\sqrt{\beta\lambda+r_{ij}^2}}$$

where ε and σ are the standard Lennard-Jones parameters, r_{ij} is the interatomic distance, q_i and q_j are the charges, ε_0 is the vacuum permittivity, and $\alpha = 0.5$ and $\beta = 12$ Å^2 the are soft-core parameters. Similarly, for all interactions with the appearing fluorine atom in **4**, the interaction energy is modified into

$$V_1(\text{appearing}) = 4\epsilon\left[\frac{1}{\left(\alpha(1-\lambda)+(r_{ij}/\sigma)^6\right)^2} - \frac{1}{\alpha(1-\lambda)+(r_{ij}/\sigma)^6}\right] + \frac{q_i q_j}{4\pi\epsilon_0\sqrt{\beta(1-\lambda)+r_{ij}^2}}$$

It should be noted that the interactions at the QM level cannot be modified in this manner, and therefore the QM and MM systems are not strictly equivalent at intermediate λ values.

NBB and its indirect version

In the general formulation of NBB, there is, for each λ, a true potential and a reference potential, with which the simulation is performed, and then $V_\lambda^{bias} = V_\lambda^{ref} - V_\lambda^{true}$. For the endpoints ($\lambda = 0$ and $\lambda = 1$), the true potential must of course correspond to the QM potential if we want to compute free energies at the QM level. However, for the intermediate λ values, we are free to choose the true potential in any manner. In particular, we can choose to define it as a linear combination of QM energies:

$$V_\lambda^{true} = V_\lambda^{QM} = (1-\lambda)\,V_0^{QM} + \lambda\,V_1^{QM} \quad ,$$

In this case, the NBB expression is given by Eq. 8 in the main part of the manuscript. This method is simply denoted NBB in Tables 2 and 4.

In the indirect version of NBB (denoted NBB indirect in Table 2), we instead define the true potential for intermediate λ values as a linear combination of MM energies, i.e. equal to the reference potential:

$$V_\lambda^{true} = \begin{cases} V_0^{QM} & \text{if } \lambda = 0 \\ V_1^{QM} & \text{if } \lambda = 1 \\ V_\lambda^{MM} = (1-\lambda)\,V_0^{MM} + \lambda\,V_1^{MM} & \text{otherwise} \end{cases}$$

In this case, Eq. 8 is replaced by the more general expression:

$$\Delta G^{A\to B} = kT\left(\ln\frac{\langle f\left(V_A^{true} - V_B^{true} + C\right)\exp\left(V_B^{bias}/kT\right)\rangle_B \langle \exp\left(V_A^{bias}/kT\right)\rangle_A}{\langle f\left(V_B^{true} - V_A^{true} + C\right)\exp\left(V_A^{bias}/kT\right)\rangle_A \langle \exp\left(V_B^{bias}/kT\right)\rangle_B}\right) + C$$

and we note that $V^{bias} = 0$ for any λ value except the two endpoints.