


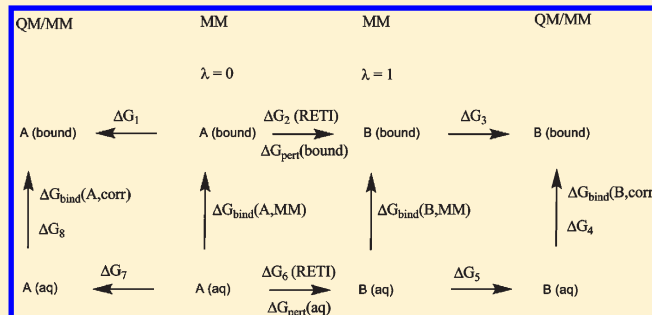
A Simple QM/MM Approach for Capturing Polarization Effects in Protein–Ligand Binding Free Energy Calculations

Frank R. Beierlein,[†] Julien Michel,[‡] and Jonathan W. Essex*

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

 Supporting Information

ABSTRACT: We present a molecular simulation protocol to compute free energies of binding, which combines a QM/MM correction term with rigorous classical free energy techniques, thereby accounting for electronic polarization effects. Relative free energies of binding are first computed using classical force fields, Monte Carlo sampling, and replica exchange thermodynamic integration. Snapshots of the configurations at the end points of the perturbation are then subjected to DFT-QM/MM single-point calculations using the B3LYP functional and a range of basis sets. The resulting quantum mechanical energies are then processed using the Zwanzig equation to give free energies incorporating electronic polarization. Our approach is conceptually simple and does not require tightly coupled QM and MM software. The method has been validated by calculating the relative free energies of hydration of methane and water and the relative free energy of binding of two inhibitors of cyclooxygenase-2. Closed thermodynamic cycles are obtained across different pathways, demonstrating the correctness of the technique, although significantly more sampling is required for the protein–ligand system. Our method offers a simple and effective way to incorporate quantum mechanical effects into computed free energies of binding.



INTRODUCTION

Computer algorithms that predict the geometry of a complex of an enzyme with a substrate or an inhibitor have become a valuable tool in structure-based drug design, both for the analysis of protein–ligand binding interactions of a single compound, or for virtual screening of large compound databases in search of new inhibitors of a pharmaceutically relevant enzyme.^{1–4} However, the prediction of the ligand binding mode is not sufficient to identify novel druglike molecules. The prediction of binding affinity, the so-called scoring problem, is essential to support structure-based drug design efforts. Commonly used scoring functions attempt to correlate experimentally determined binding affinities with calculated properties of the protein–ligand complex. Scoring functions are designed to be relatively efficient as they need to process hundreds of thousands of docking poses in virtual screening experiments and therefore often sacrifice accuracy for speed.^{4,5} Properly parametrized scoring functions have been shown to discriminate to some degree compounds that bind to a target from those which do not, but their performance is often target dependent. In addition, scoring functions often have difficulties in properly rank-ordering active compounds by their potency.^{4,5}

In contrast to common scoring functions, rigorous free energy calculations allow, in principle, an exact computation of the binding free energy of a ligand.^{6,7} This approach has been in use for more than two decades,⁸ and there have been many improvements and renewed interest in the application of rigorous free

energy calculations to drug design problems recently.^{6,9–17} Even data sets of structurally diverse compounds can now be handled (see, e.g., refs 13 and 14 and the references cited therein). In practice, however, these statistical mechanics equations are far too complex to be solved rigorously even with the aid of a computer and approximate solutions must be employed. For example, molecular interactions are often modeled by molecular force fields rather than quantum theory.¹⁸ The accuracy of the predicted free energy will be reduced if the Hamiltonian used does not reproduce well the intermolecular interactions. For example, polarization of the ligand by the environment (protein or solvent) is normally not considered explicitly, as nonpolarizable force fields are used in most studies. In some cases, polarization of a solute by the solvent is implicitly considered by using solute charges derived by a method that overestimates the solute gas-phase charges to some extent, e.g., RESP 6-31G* charges in the case of the Amber 94/99/Gaff series of force fields.^{19–21}

Although linear-scaling QM approaches^{22–29} in principle allow one to treat a macromolecule completely as a QM system, hybrid quantum mechanical/molecular mechanical (QM/MM) approaches³⁰ are still standard techniques for the description of large molecular systems, such as enzymes (see, e.g., refs 31–37 and the references cited therein). The fundamental concepts of

Received: September 22, 2010

Revised: January 28, 2011

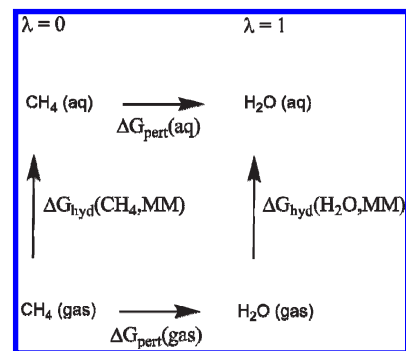
Published: April 08, 2011

QM/MM techniques were given by Warshel and Levitt in 1976.³⁰ The basic idea of all QM/MM approaches is to treat that part of the system that undergoes the most important electronic changes during a reaction, or upon binding a substrate, quantum mechanically, and the rest of the system by molecular mechanics.³⁰

While the focus of the current work is on the calculation of relative protein–ligand binding free energies and solvation free energies, this problem is conceptually related to the calculation of reaction free energies (potentials of mean force, PMF). A brief overview of the literature for both problems is given in the Supporting Information. The computational costs for extensively sampling a solvated protein–ligand system with a QM or QM/MM Hamiltonian are (and will be for a while) prohibitive, for use in the context of binding free energy calculations. A promising approach to circumvent this problem is to use a fast but less accurate method to sample phase space and to subsequently correct the simulation results using a modest number of high-level QM/MM energy calculations.^{38–64} This approach was, for example, used to calculate reaction free energies for chemical reactions in enzymes^{39–41,55,58} and in solution^{54,57,59} as well as solvation free energies of ions and small molecules^{42,44,45,47,60,61} using a relatively cheap reference potential like molecular mechanics (MM), or, in some cases, an empirical valence bond (EVB)^{55,57–59} potential. Correction free energies were obtained by using a free energy perturbation (FEP),^{39–47,59,60,62,63} a thermodynamic integration (TI),^{38,64} or a linear response approximation (LRA)^{55,57} approach. Another interesting idea is to generate a rigorous QM/MM ensemble from an MM simulation by periodically performing QM/MM calculations on MM geometries and accepting/rejecting the last set of configurations by application of a Metropolis test (Metropolis–Hastings⁶⁵ approach).^{38,48–51,64}

Most of the approaches that use a fast reference potential in combination with QM/MM processing of a subset of configurations have been used to calculate reaction free energies or solvation free energies of small molecules; only very few implementations have addressed the problem of calculating binding free energies of pharmacologically relevant protein–inhibitor systems. Therefore, we have developed a simple, easy to implement approach to consider electronic polarization effects in rigorous calculations of protein–inhibitor binding free energies. We first employ a standard thermodynamic cycle approach, which uses standard Monte Carlo replica exchange simulations to ensure sampling of the MM free energy landscape and provides free energies differences between pairs of compounds using thermodynamic integration.^{9,16,17} Subsequently, we add additional legs to the thermodynamic cycle and use the classical configurations to compute the free energy difference between a classical MM and a QM/MM representation of the system, similar to the approaches used by Warshel^{52,54,60,61} and Wood.^{42–47} Our approach corresponds to a free energy perturbation (FEP) between MM and QM/MM potentials, without intermediate states. As our current implementation does not permit the sampling of the QM/MM potential surface (for reasons of computational expediency), we have designed a novel test to check the pathway independence of the free energies obtained. A further consequence of not sampling the QM/MM potential surface is that we have not been able to use a linear response approximation approach as suggested in refs 55 and 57 or other techniques that require QM/MM sampling, e.g., that described in ref 38. We are aware that the best representation of a system would be all-QM, or at least some approximation that allows quasi-QM sampling (e.g., the schemes used by Iftimie et al.^{49,50} and Woods et al.³⁸), but for reasons of computational

Scheme 1. Thermodynamic Cycle Used in the Pure MM Simulations of the Validation System (Methane to TIP4P Water) To Obtain Relative Solvation Free Energies of Two Compounds



expediency, both in terms of computational cost, and the availability of the appropriate software, we have chosen a different approach.

The combination of a standard, nonpolarizable force field with QM/MM to incorporate polarization effects allows an accurate description of electronic polarization effects in the different legs of the thermodynamic cycles (subject to the level of QM theory chosen), using comparatively low-quality MM parameters. We believe that this is a universally applicable approach that is able to treat electronic polarization in a wide range of different cases, e.g., in (aqueous) solution and in different kinds of binding pockets.

The aim of our current work is not the exact reproduction of experimental binding/solvation free energies but to investigate whether it is possible to obtain reproducible, converged, free energy results using a very simple and computationally inexpensive approach, without the requirement for dedicated, closely coupled QM and MM programs. We also introduce a novel test to check the convergence and numerical stability of our MM \rightarrow QM/MM FEP approach. Further studies will need to examine the accuracy of the approach by investigating the compatibility of the Hamiltonians used for MM and QM/MM, an issue that has gained considerable interest recently.⁶⁴

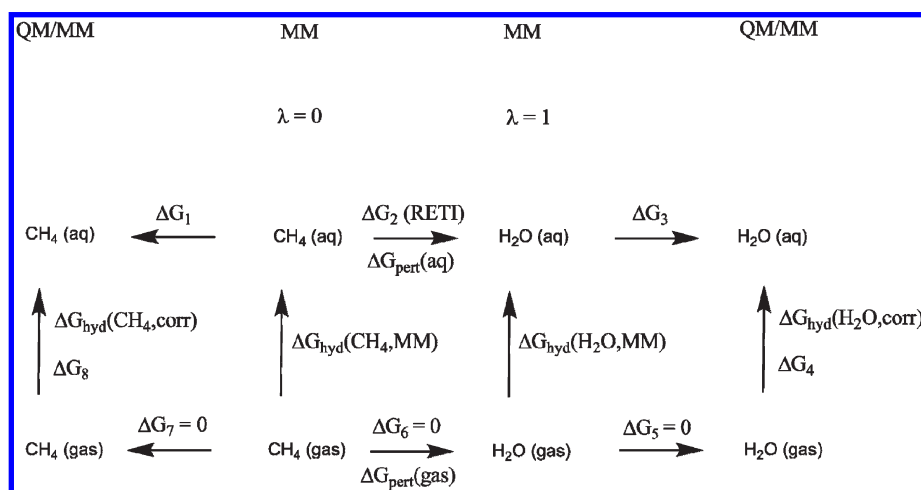
■ VALIDATION SYSTEM: METHANE TO TIP4P WATER

Method: MM Free Energy Calculations. To validate our approach to include polarization in free energy thermodynamic integration studies, we focus first on a simple test system: the perturbation of methane to water, performed twice, once in water as solvent, and in the gas phase. Using a standard thermodynamic cycle as depicted in Scheme 1 allows the computation of the relative free energies of solvation of the two compounds.

In the MM step we use a standard Monte Carlo (MC) thermodynamic integration simulation, which uses a replica exchange approach to improve sampling of the free energy surface (replica exchange thermodynamic integration, RETI).^{16,17}

Computational Details: MM Free Energy Calculations. Methane was perturbed into TIP4P⁶⁶ water using a standard single topology perturbation approach. The solute (methane and water, respectively) was kept rigid, and only solvent molecules were allowed to move. The solvent was represented by a periodic box of 905 TIP4P water molecules with dimensions of about 30 Å. Simulations were performed in the NPT ensemble at a temperature of 298.15 K and a pressure of 1 atm. GAFF²¹ parameters

Scheme 2. Thermodynamic Cycle with QM/MM Corrections Used for the Calculation of Relative Free Energies of Solvation of Methane and Water



together with AM1-BCC^{67,68} charges were used for all-atom CH₄, TIP4P⁶⁶ parameters for H₂O. For the nonbonded interactions a residue-based cutoff of 10 Å was used. No intramolecular interactions spanning more than two bonds were present in the system, and thus, according to the force field used, all intramolecular nonbonded interactions were neglected. Dummy atoms were retracted into the van der Waals radii of the neighboring atoms as they were turned off to avoid end point singularity problems. To ensure a smooth transition between methane and water, 12 almost equally spaced λ windows (0.00, 0.09, 0.18, 0.27, 0.36, 0.45, 0.54, 0.64, 0.73, 0.82, 0.91, 1.00) were used for the free energy calculation.

The Amber 9 suite⁶⁹ was used for parameter generation. Monte Carlo simulations were performed using ProtoMS 2.2.⁷⁰ Replica exchange thermodynamic integration (RET) was used as described previously.^{16,17} 10 million equilibration moves were performed at each λ window, and five independent runs of 40 million moves were used for data collection.

Results: MM Free Energy Calculations. The simulations provided smooth gradients (not shown) of $dG/d\lambda$ which were used for thermodynamic integration. Relative free energies of hydration were calculated according to Scheme 1/eq 1:

$$\Delta G_{\text{hyd}}(\text{H}_2\text{O}, \text{MM}) - \Delta G_{\text{hyd}}(\text{CH}_4, \text{MM}) = \Delta G_{\text{pert}}(\text{aq}) - \Delta G_{\text{pert}}(\text{gas}) \quad (1)$$

As the solute was kept rigid, the thermodynamic cycle shown in Scheme 1 is simplified because $\Delta G_{\text{pert}}(\text{gas})$ is zero. We obtain a value for $\Delta\Delta G_{\text{hyd}} = \Delta G_{\text{hyd}}(\text{H}_2\text{O}) - \Delta G_{\text{hyd}}(\text{CH}_4)$ of -8.88 ± 0.06 kcal·mol⁻¹ (one standard error). This is in reasonable agreement with the experimental value,⁷¹ which is -8.31 kcal·mol⁻¹, and with previous results by Woods et al. (-8.9 ± 0.1 kcal·mol⁻¹).³⁸

Method: QM/MM Corrections. To provide a polarization correction to the free energies of solvation obtained classically, we add additional legs to the thermodynamic cycle used for the classical MM free energy calculations shown above (Scheme 2). Configurations at the end points ($\lambda = 0$ and $\lambda = 1$) of the classical free energy simulation are selected and used as inputs for DFT-QM/MM single-point calculations. Free energies that reflect the difference between an MM and a QM/MM description of the system are calculated with the Zwanzig equation,⁷² using the difference of the solute/solvent interaction energies in the QM/MM and the MM case, respectively, in the exponential term of the

equation (eq 2). Thus, we are pursuing a free energy perturbation (FEP) approach to describe the difference between an MM and a QM/MM representation of the system.

$$\Delta G_{\text{MM} \rightarrow \text{QM/MM}} = -RT \ln \left\langle \exp \left[- (U_{\text{QM/MM}} - U_{\text{QM,vac}} - U_{\text{charges,MM}} - U_{\text{Coul,solute-solv,MM}}) / RT \right] \right\rangle \quad (2)$$

$U_{\text{QM/MM}}$ is the total energy as obtained from a QM/MM calculation with background charges that polarize the QM wave function (the solute serves as QM part, the background water charges as MM part of the QM/MM system). $U_{\text{QM,vac}}$ is the single-point energy of the solute (the QM part) in vacuum, and $U_{\text{charges,MM}}$ is the sum of all Coulomb interactions of the background charges. $U_{\text{Coul,solute-solv,MM}}$ is the Coulomb solute-solvent interaction energy as obtained from the MM part of the protocol. Therefore, we subtract the solute-solvent Coulomb energy for an MM treatment of the system from the corresponding solute-solvent interaction energy as obtained from QM/MM.

Unlike standard FEP implementations, we do not use intermediate states between MM and QM/MM, and we only sample phase space using the MM Hamiltonian. Normally, we would also sample phase space using the QM/MM Hamiltonian, and would calculate $\Delta G_{\text{QM/MM} \rightarrow \text{MM}}$, which should be similar in value to $-\Delta G_{\text{MM} \rightarrow \text{QM/MM}}$ in the limit of negligible hysteresis. Therefore, our approach is an approximation which uses a simple post processing by selecting a number of MM configurations for QM/MM single-point calculations. We later demonstrate that such a crude approximation is acceptable for the systems studied in this work. Standard codes are used to perform MM and QM/MM calculations and these two levels of theory are linked by simple Perl⁷³ scripts.

Lastly, we are using a simple point charge model in the QM/MM calculations. This means that, in addition to polarization of the QM part by the MM charges, only Coulomb interactions between the MM environment and the QM part of the system are corrected. van der Waals interactions between solute and solvent are captured by the MM part of our protocol.

QM/MM Calculations: Computational Details. Configurations from the “end points” of the MM simulation ($\lambda = 0$ and $\lambda = 1$,

Table 1. Free Energies and Standard Errors As Defined in Scheme 2 for the System Methane–Water (kcal·mol^{−1})^a

	B3LYP				
	6-31G*	6-31G**	cc-pVTZ	cc-pVQZ	AUG-cc-pVQZ
$\Delta G_1(\text{CH}_4)$	−0.17 ± 0.01	−0.17 ± 0.01	−0.20 ± 0.02	−0.23 ± 0.02	−0.30 ± 0.02
$\Delta G_2(\text{RETI})$	−8.88 ± 0.06	−8.88 ± 0.06	−8.88 ± 0.06	−8.88 ± 0.06	−8.88 ± 0.06
$\Delta G_3(\text{H}_2\text{O})$	−2.03 ± 0.04	−1.55 ± 0.03	−1.51 ± 0.05	−1.64 ± 0.06	−1.61 ± 0.06
$\Delta G_4 - \Delta G_8$	−10.74 ± 0.08	−10.26 ± 0.07	−10.19 ± 0.08	−10.29 ± 0.09	−10.18 ± 0.09

^a Standard errors were calculated from batch averages over five independent runs.

which correspond to “real” methane and water) were selected for subsequent QM/MM single-point calculations with Gaussian 03.⁷⁴ One QM single-point calculation of the solute with background charges representing the solvent (Gaussian keyword “CHARGE”) was performed every 100 000 MM Monte Carlo moves. This gave a total of 2000 QM/MM single points for methane and water each. Only charges of solvent molecules that are within the cutoff used in the MM part of the protocol (see above) were considered for the QM/MM calculations. Gaussian 03 calculations with background charges allow a polarization of the QM wave function by the MM charges; no back-polarization of the MM part by the QM part was considered.

The QM energies were computed using B3LYP⁷⁵ hybrid density functional calculations with different basis sets, as implemented in Gaussian 03:⁷⁴ 6-31G*, 6-31G**,^{76–85} cc-pVTZ, cc-pVQZ, and AUG-cc-pVQZ.^{86–90}

The free energies shown in Scheme 2 are calculated as follows: the MM free energy change ΔG_2 is calculated by standard replica exchange thermodynamic integration (RETI). The gas-phase free energy ΔG_6 is zero because methane and water are treated as rigid molecules. ΔG_1 and ΔG_3 account for the free energy change upon switching from an MM to a QM/MM representation of solute–solvent electrostatic interactions. They were calculated using the formulation of the Zwanzig⁷² equation shown above (eq 2). ΔG_5 and ΔG_7 are also zero because methane and water do not have intermolecular interactions in the gas phase.

Since the free energy is a state function, from Scheme 2 we have

$$-\Delta G_1 + \Delta G_2 + \Delta G_3 - \Delta G_4 - \Delta G_5 - \Delta G_6 + \Delta G_7 + \Delta G_8 = 0 \quad (3)$$

giving an expression for the QM/MM-corrected difference in the free energies of hydration of water and methane (incorporating the zero values for the gas-phase free energies ΔG_5 , ΔG_6 , ΔG_7):

$$\Delta G_{\text{hyd}}(\text{H}_2\text{O}, \text{corr}) - \Delta G_{\text{hyd}}(\text{CH}_4, \text{corr}) = \Delta G_4 - \Delta G_8 = -\Delta G_1 + \Delta G_2 + \Delta G_3 \quad (4)$$

Results and Discussion: Solvation Free Energies of Methane and Water, with QM/MM Corrections. The free energies for the perturbation methane–TIP4P water as defined in Scheme 2 are summarized for several combinations of the B3LYP functional and different basis sets in Table 1. The QM/MM-corrected relative hydration free energies $\Delta G_{\text{hyd}}(\text{H}_2\text{O}, \text{corr}) - \Delta G_{\text{hyd}}(\text{CH}_4, \text{corr})$ shown in Table 1 are in the range of −10.7 to −10.2 kcal·mol^{−1}; all are too negative. For the range of basis sets investigated in our study, the best results are achieved with the computationally less demanding basis sets 6-31G** and cc-pVTZ. It is interesting to note that previous studies have also made the observation that QM/MM corrections lead to a decrease in accuracy

compared to experiment, e.g. in the studies by Woods et al.³⁸ and Wesolowski and Warshel,⁶⁰ who used different approaches and QM models.

ΔG_1 , which accounts for the free energy difference between an MM and a QM/MM potential for CH₄, is in the range of −0.30 to −0.17 kcal·mol^{−1}, which indicates that, at this level of QM theory, the electronic polarization effects are reasonably well modeled by the AM1-BCC charges used in the MM calculation.

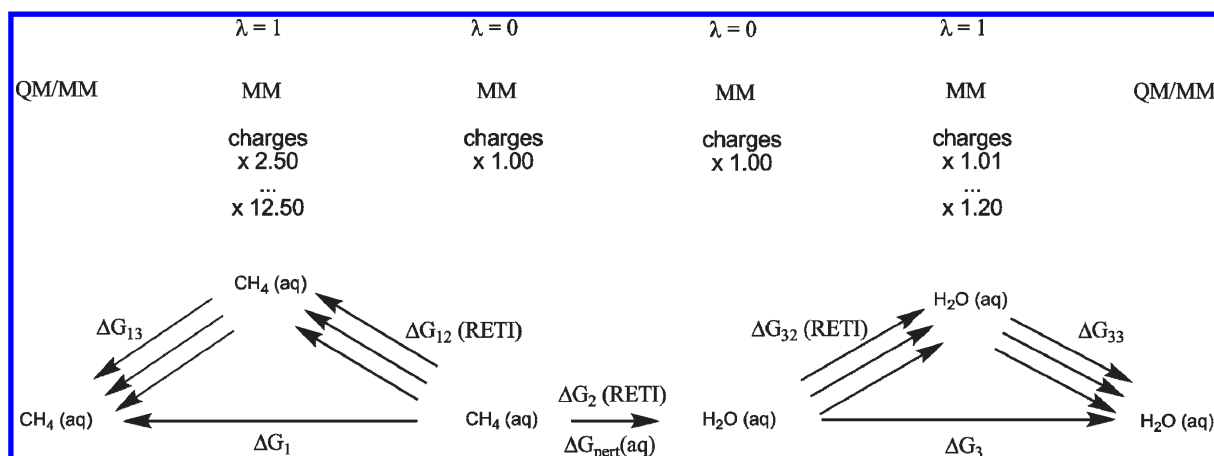
ΔG_3 (MM→QM/MM for H₂O) varies in the range of −2.0 to −1.5 kcal·mol^{−1}, which is a much larger energy than in the methane case; similar numbers have been observed in the aforementioned previous study by Woods et al.³⁸ This indicates that, within the limits and approximations imposed by the QM model, the classical water model is underpolarized and that polarization by the bulk solvent decreases the hydration free energy. However, systematic improvements in the description of molecular energetics should not be expected: sampling of the free energy landscape is only performed using the MM Hamiltonian, only polarization of the QM part by the MM environment is considered, and there is no back-polarization of the MM environment by the QM part. Finally, the TIP4P water model is already highly optimized to model condensed phase properties and therefore does not actually need a correction as performed in this study. The methane to water model was merely selected as a test case to show the general applicability of such an approach.

Method: “Charge Perturbations” (Thermodynamic Cycle Closure). In a normal FEP study, it would be standard practice to compute forward (MM→QM/MM) and backward (QM/MM→MM) free energies to assess the convergence of the computed free energy changes. As our implementation does not allow the sampling of configurations under the QM/MM Hamiltonian (because this would be too computationally expensive), we need to develop another way to prove the validity of our approach.

As the Gibbs free energy is a thermodynamic state function, its value at a given state is independent of the pathway that was used to reach this state. To prove that a calculated free energy is converged, it is therefore possible to generate alternative pathways to reach this state. If all pathways lead to the same overall calculated free energy, then it is likely, although not guaranteed, that the approach is correct.

To construct a series of alternative pathways, we perform classical MM RETI perturbations from the solute with standard AM1-BCC charges to the solute with charges scaled by an arbitrary factor. These alternative configurations, which do not necessarily reflect chemically reasonable species, are then used as input structures for DFT-QM/MM single-point calculations, and an FEP (MM→QM/MM) free energy is calculated using eq 2. The sum of the MM RETI charge perturbation free energy and the FEP (MM→QM/MM) free energy must be equal to the “normal” FEP (MM→QM/MM) free energy (i.e., the one

Scheme 3. Thermodynamic Cycle with Charge Perturbations Used To Show the Pathway Independence of the Relative Free Energies of Hydration of Methane and Water^a



^a Alternative configurations are generated by scaling the MM force field charges by a range of factors and applying the QM/MM corrections to the configurations generated by this alternative force field.

without charge perturbations) to obtain closed thermodynamic cycles (Scheme 3).

Computational Details: “Charge Perturbations” (Thermodynamic Cycle Closure). Scheme 3 defines the alternative pathways used to investigate the pathway independence of the MM \rightarrow QM/MM free energies for the methane to water system. MM RETI simulations (“charge perturbations”) were performed according to a similar protocol as explained above for the “normal” perturbations of methane into TIP4P water. Ten million equilibration moves were performed at each λ window, and 40 million moves were used for data collection. TIP4P water with original charges ($\lambda = 0$) was perturbed into water with charges scaled by a factor (1.01, 1.05, 1.07, 1.10, 1.15, 1.20, $\lambda = 1$). Six evenly spaced λ windows (0.00, 0.20, 0.40, 0.60, 0.80, 1.00) were used for the charge perturbations. Analogous simulations were performed for methane (factors 2.50, 5.00, 7.50, 8.50, 10.00, 12.50). QM/MM postprocessing of the configurations with the scaled charges was done analogously to the noncharge scaled calculations, using B3LYP/6-31G* and B3LYP/cc-pVTZ. For the B3LYP/6-31G* calculations, additional charge scale factors were investigated (water 1.01, 1.05, 1.07, 1.08, 1.09, 1.10, 1.15, 1.20; methane 2.50, 5.00, 5.26, 7.50, 8.00, 8.25, 8.50, 10.00, 12.50).

The free energies that describe the MM charge perturbations, ΔG_{32} and ΔG_{12} , are obtained from classical RETI simulations using ProtoMS 2.2.⁷⁰ ΔG_{33} (water) and ΔG_{13} (methane) are calculated using eq 2 and represent FEP energies that describe the difference in the solute–solvent Coulombic interaction free energy between QM/MM and MM because of polarization.

The pathway independence of the free energy requires that for all pathways investigated

$$\Delta G_{32} + \Delta G_{33} = \Delta G_3 \text{ (water)} \quad (5)$$

$$\Delta G_{12} + \Delta G_{13} = \Delta G_1 \text{ (methane)} \quad (6)$$

Charge Perturbations: Results and Discussion. RETI charge perturbations for methane and water provided smooth gradients (see Supporting Information, Figure S1), which were integrated to give ΔG_{32} (water) and ΔG_{12} (methane). Figures 1 and 2 show these free energies as well as the MM \rightarrow QM/MM FEP free

energies ΔG_{33} and ΔG_{13} for different charge scale factors for the B3LYP calculations using the cc-pVTZ basis set. B3LYP/6-31G* results are qualitatively similar and are provided in the Supporting Information (Figures S2 and S3). The sums of the RETI energies ΔG_{32} and ΔG_{12} and their corresponding MM \rightarrow QM/MM FEP energies ΔG_{33} and ΔG_{13} are constant for different charge scale factors, indicated by the small slope of the corresponding linear regression. Interestingly, for all alternative pathways investigated, the mean values of the sums $\Delta G_{12} + \Delta G_{13}$ and $\Delta G_{32} + \Delta G_{33}$ are in good agreement with ΔG_1 and ΔG_3 , respectively (Table S1 in the Supporting Information). Given that the same QM/MM-corrected free energy of hydration is obtained, irrespective of the charge set used, it can be concluded that, at least for this system, our approximation of performing FEP between MM and QM/MM in one step is a viable approach.

In principle, charge scaling could be used to obtain “ideal” polarized charges of a solute in a polarizing environment, at least in the case of an isotropically polarizing medium—the ideal charge scale factor is the one at which the QM/MM-FEP correction free energy is zero. This interpretation is probably acceptable for a solvent environment, but certainly not for a protein binding pocket, which has typically very different properties in different regions of the active site. For water, our calculations show that a scale factor of ~ 1.07 – 1.08 (vs TIP4P) is required to capture the polarizing effect of a water environment, modeled using this QM/MM formalism. For methane, a scale factor of ~ 8.0 – 8.5 vs the charges obtained by the AM1-BCC charge fitting method is required to model the “real” (polarized) behavior of methane, under the approximations of our QM/MM model. These large factors can be understood if one recalls that the first nonzero multipole moment of methane is an octopole. Interestingly, the octopole moment is sometimes modeled in point-charge models by using much larger (factor 5.26, C -0.572 , H 0.143) partial charges than those derived from the AM1-BCC method (C -0.10872 , H 0.02718).^{91,92}

Evaluation System (Methane to Water): Summary and Outlook. The aim of our current study was the development of a simple protocol that uses existing codes to obtain a QM/MM correction to free energies of solvation/binding calculated by a pure MM approach. As a test case, we used the perturbation of

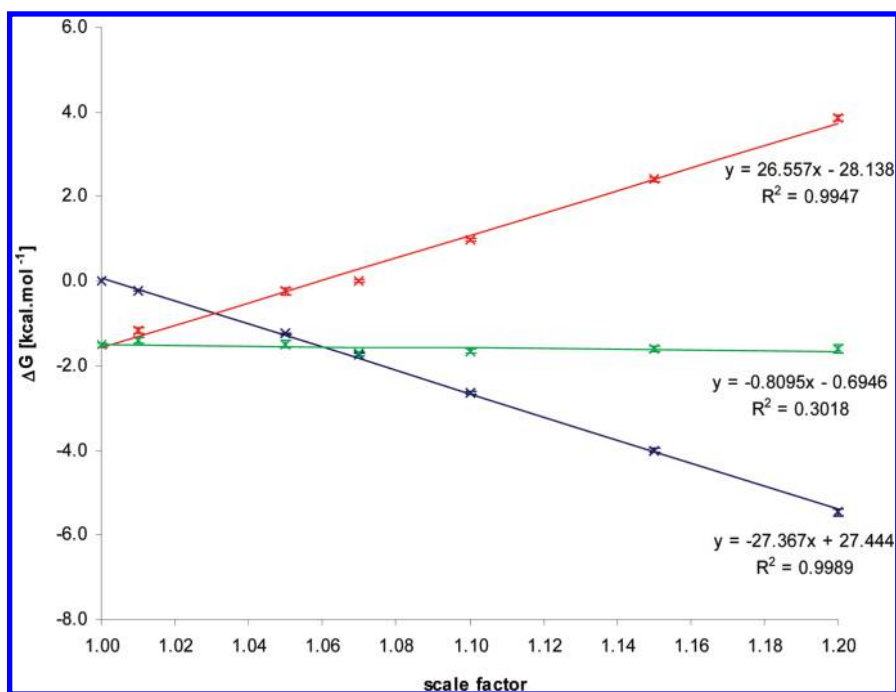


Figure 1. Free energies obtained from charge perturbations for water (B3LYP/cc-pVTZ). Blue: classical RETI energies (ΔG_{32}). Red: QM/MM-FEP energies calculated by eq 2 (ΔG_{33}). Green: sum of classical and QM/MM-FEP energies (thermodynamic cycle closure). Mean value over all charge scale factors $\langle \Delta G_{32} + \Delta G_{33} \rangle = -1.58 \pm 0.04 \text{ kcal} \cdot \text{mol}^{-1}$; $\Delta G_3 -1.51 \pm 0.05 \text{ kcal} \cdot \text{mol}^{-1}$.

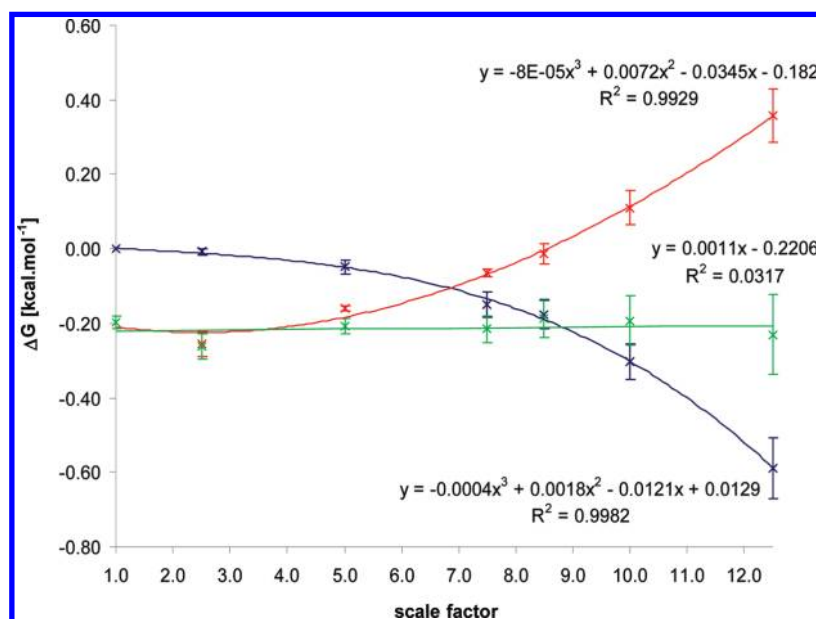


Figure 2. Free energies obtained from charge perturbations for methane (B3LYP/cc-pVTZ). Blue: classical RETI energies (ΔG_{12}). Red: QM/MM-FEP energies calculated by eq 2 (ΔG_{13}). Green: sum of classical and QM/MM-FEP energies (thermodynamic cycle closure). Mean value over all charge scale factors $\langle \Delta G_{12} + \Delta G_{13} \rangle = -0.22 \pm 0.01 \text{ kcal} \cdot \text{mol}^{-1}$; $\Delta G_1 -0.20 \pm 0.02 \text{ kcal} \cdot \text{mol}^{-1}$.

methane to water in water. Closed thermodynamic cycles prove the pathway independence and thus validate our approach, at least for the simple test system chosen.

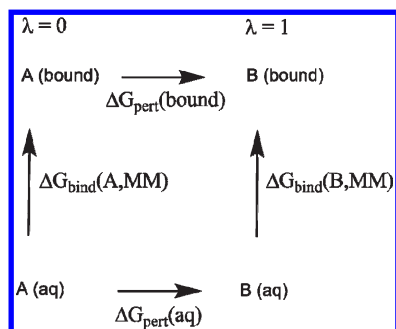
It was not the intention to calculate accurate free energies of solvation for methane/water in water. These free energies were obtained in sufficient quality by the MM part of the protocol and the QM/MM-FEP correction did not improve the actual numbers in this case. This can be understood as the MM parameters

are, at least in the case of TIP4P water, highly optimized and thus the MM free energy is already in good agreement with experiment.

APPLICATION TO A PROTEIN–LIGAND SYSTEM: COX-2

While our simple approach of performing FEP between an MM and a QM/MM representation of the system of interest works

Scheme 4. Thermodynamic Cycle for the Calculation of Relative Protein–Ligand Binding Free Energies Using a Classical MM Approach



well in calculating relative solvation free energies of rigid methane and TIP4P water, it is interesting to apply the approach to a much more challenging simulation system, e.g., the calculation of relative binding free energies of two inhibitors of an enzyme. In this case, it is much more difficult to achieve adequate sampling of the energy landscapes, because the ligands and parts of the protein are treated as flexible in the simulation. A standard thermodynamic cycle used to calculate the relative binding free energies of two protein inhibitors is depicted in Scheme 4.

Relative binding free energies are calculated according to eq 7.

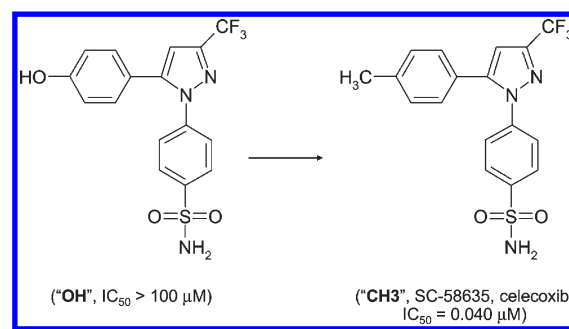
$$\Delta G_{\text{bind}}(\text{B, MM}) - \Delta G_{\text{bind}}(\text{A, MM}) = \Delta G_{\text{pert}}(\text{bound}) - \Delta G_{\text{pert}}(\text{aq}) \quad (7)$$

Cyclooxygenase-2 has been chosen as an example to calculate relative protein–ligand binding affinities as it has been extensively studied in pure MM studies previously.^{9,13,93} For the current study, we have chosen to investigate the perturbation of a ligand bearing a hydroxylic group (ligand “OH” in Scheme 5) into the methyl derivative (ligand “CH₃” in Scheme 5), for which an experimental relative binding free energy $\Delta\Delta G < -4.64$ kcal·mol^{−1} has been reported.⁹⁴ Previous RETI simulations in our lab (with Gaff parameters and AM1-BCC charges) have determined relative binding free energies of -2.99 ± 0.07 kcal·mol^{−1} (ref 13) and -2.9 ± 0.37 kcal·mol^{−1}.⁹

The perturbation described above seems to be ideal to investigate the applicability of our MM → QM/MM FEP approach as the classical RETI energy is too low^{9,13} compared with the experimental results. Additionally, there seems to be a large influence of the charges used in the classical simulations. While the perturbations performed in our lab (RETI, Gaff/AM1-BCC) yielded a relative binding free energy which is approximately 1.6 kcal·mol^{−1} too positive, Plout Price and Jorgensen report a value of -5.39 ± 0.21 kcal·mol^{−1} with the OPLS force field and CM1A charges scaled by 1.2.⁹³

While it has been shown that the quality of the charges used in an MM perturbation study is essential for the accuracy of the results,^{95–97} our approach in this work is to use inexpensive and well-established atomic partial charges like AM1-BCC, and to postprocess the configurations using our QM/MM approach. Doing so has not only the benefit that it provides a general solution to the problem of deriving force field parameters for organic molecules at a moderate cost but also that the polarization arising from different regions of the protein can be captured directly.

Scheme 5. Perturbation Applied to Transform Cox-2 Inhibitor OH into CH₃ (Celecoxib)



Computational Details: MM Free Energy Calculations (Protein–Ligand System). In the MM part of the protocol, replica exchange thermodynamic integration (RETI)^{16,17} simulations were performed according to a similar protocol as published previously.^{9,13} Full details are given in the Supporting Information. A single topology approach was used to perturb ligand OH into ligand CH₃. The solutes were treated flexibly (with the exception of rings); protein side chains within a radius of 10 Å from any solute atom were also treated flexibly. After initial equilibration of the system, each λ window (0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 1.00) was equilibrated for a further 10 million moves and data were collected during 30 million simulation moves consisting of 300 blocks with 100 000 Monte Carlo moves each, which were then repeated 10 times (10 runs). Two different sampling schemes were compared: first, 10 completely independent runs were performed, each starting from the equilibrated X-ray structure, with a new random seed for each run (“repeated sampling”). Then, another series of 10 runs was performed in which the last configuration of the previous run was used as starting structure for the next (“extended sampling”), also with a new random seed for each run. The RETI energies reported below (Tables 2 and 3) are mean values of 10 runs.

Computational Details: QM/MM Calculations (Protein–Ligand System). Only the “end points”, i.e., trajectories for $\lambda = 0$ and $\lambda = 1$ of the RETI simulations explained above, were used for QM/MM postprocessing. Snapshots for QM/MM were selected at the end of each MM simulation block, i.e., after 100 000 MM Monte Carlo moves. Therefore, 3000 QM/MM single-point calculations had to be performed for $\lambda = 0$ and the same number for $\lambda = 1$ (300 blocks × 10 calculations). Gaussian 03⁷⁴ single-point calculations were performed for the solute, and the solvent/protein was represented by background point charges (G03 CHARGE keyword). Only charges of solvent molecules/protein residues which fulfill the cutoff criterion used by ProtoMS⁷⁰ were used for QM/MM; they were scaled in the 0.5 Å feather region. B3LYP⁷⁵ hybrid density functional calculations were performed with the 6-31G**^{76–85} basis set.

Scheme 6 shows the thermodynamic cycle used to calculate QM/MM-corrected protein–ligand binding free energies. ΔG_2 and ΔG_6 are obtained from classical MM RETI simulations as explained above. ΔG_1 and ΔG_3 describe the difference in the solute–protein/solvent interaction between QM/MM and pure MM because of polarization and are calculated according to eq 8. ΔG_7 and ΔG_5 are the analogous free energies for the unbound simulation leg (solute–solvent

Table 2. Free Energies As Defined in Scheme 6 for Complexes of Ligands OH and CH₃ with Cox-2 (“Repeated Sampling”, 10 Completely Independent Runs, Each Starting from the Equilibrated X-ray Structure)^a

			ΔG	$\langle \Delta U \rangle$	% ($\Delta U < 0$)
bound	ΔG_1	QM/MM-FEP OH	0.28 ± 0.23	2.64 ± 0.40	4.8
	ΔG_2	RETI OH \rightarrow CH ₃	15.06 ± 0.04		
	ΔG_3	QM/MM-FEP CH ₃	-0.64 ± 0.28	2.19 ± 0.48	9.3
free	ΔG_5	QM/MM-FEP CH ₃	-0.71 ± 0.26	3.57 ± 0.62	6.4
	ΔG_6	RETI OH \rightarrow CH ₃	18.14 ± 0.07		
	ΔG_7	QM/MM-FEP OH	-1.18 ± 0.27	2.77 ± 0.96	12.9
	$\Delta G_4 - \Delta G_8$	$= -\Delta G_1 + \Delta G_2 + \Delta G_3 - \Delta G_5 - \Delta G_6 + \Delta G_7$	-4.47 ± 0.53		

^a All energies are given in kcal·mol⁻¹; errors represent one standard error and were calculated from batch averages over the 10 independent runs. ΔU is the mean difference of QM/MM and MM solute–environment interaction energies ($U_{\text{QM/MM}} - U_{\text{QM,vac}} - U_{\text{charges,MM}} - U_{\text{Coul,solute-solv/protein,MM}}$ in eq 8). % ($\Delta U < 0$) is the percentage of configurations where $\Delta U < 0$, i.e., where the QM/MM interaction is stronger than the MM interaction.

Table 3. Free Energies As Defined in Scheme 6 for Complexes of Ligands OH and CH₃ with Cox-2, Obtained with an Alternative Sampling Scheme (“Extended Sampling”, 10 Runs, the Last Restart File from the Previous Run Was Used as Initial Configuration)^a

			ΔG	$\langle \Delta U \rangle$	% ($\Delta U < 0$)
bound	ΔG_1	QM/MM-FEP OH	1.74 ± 0.35	3.93 ± 1.32	1.0
	ΔG_2	RETI OH \rightarrow CH ₃	15.27 ± 0.08		
	ΔG_3	QM/MM-FEP CH ₃	1.02 ± 0.41	3.42 ± 1.32	2.2
free	ΔG_5	QM/MM-FEP CH ₃	-1.04 ± 0.28	3.31 ± 0.57	6.9
	ΔG_6	RETI OH \rightarrow CH ₃	18.14 ± 0.08		
	ΔG_7	QM/MM-FEP OH	-1.89 ± 0.18	2.09 ± 0.86	20.0
	$\Delta G_4 - \Delta G_8$	$= -\Delta G_1 + \Delta G_2 + \Delta G_3 - \Delta G_5 - \Delta G_6 + \Delta G_7$	-4.44 ± 0.64		

^a All energies are given in kcal·mol⁻¹; errors represent one standard error and were calculated from batch averages over the 10 runs. ΔU is the mean difference of QM/MM and MM solute–environment interaction energies ($U_{\text{QM/MM}} - U_{\text{QM,vac}} - U_{\text{charges,MM}} - U_{\text{Coul,solute-solv/protein,MM}}$ in eq 8). % ($\Delta U < 0$) is the percentage of configurations where $\Delta U < 0$, i.e., where the QM/MM interaction is stronger than the MM interaction.

interaction).

$$\Delta G_{\text{MM} \rightarrow \text{QM/MM}} = -RT \ln \left\langle \exp \left[- (U_{\text{QM/MM}} - U_{\text{QM,vac}} - U_{\text{charges,MM}} - U_{\text{Coul,solute-solv/protein,MM}}) / RT \right] \right\rangle \quad (8)$$

$U_{\text{QM/MM}}$ is the total energy obtained from a QM/MM calculation with background charges that polarize the QM wave function (the solute serves as QM part, the background charges as MM part of the QM/MM system). $U_{\text{QM,vac}}$ is the single-point energy of the solute (the QM part) in vacuum without background charges (one single-point calculation for each configuration), and $U_{\text{charges,MM}}$ is the sum of all Coulomb interactions of the background charges. $U_{\text{Coul,solute-solv/protein,MM}}$ is the sum of the Coulomb solute–solvent and solute–protein interaction energies as obtained from the MM part of the protocol. Thus, we use the difference of the solute–environment electrostatic interaction energy for an MM treatment of the system from the corresponding solute–environment electrostatic interaction energy in the QM/MM case in the exponential term of the Zwanzig equation.

A QM/MM-corrected relative protein–ligand binding free energy can be calculated as follows (eq 9).

$$\Delta G_{\text{bind}}(\text{B, corr}) - \Delta G_{\text{bind}}(\text{A, corr}) = \Delta G_4 - \Delta G_8 = -\Delta G_1 + \Delta G_2 + \Delta G_3 - \Delta G_5 - \Delta G_6 + \Delta G_7 \quad (9)$$

Cox-2 Results and Discussion: Free Energies, RETI and FEP (MM \rightarrow QM/MM). As expected, the results of the MM RETI simulations are in excellent agreement with previous results. The

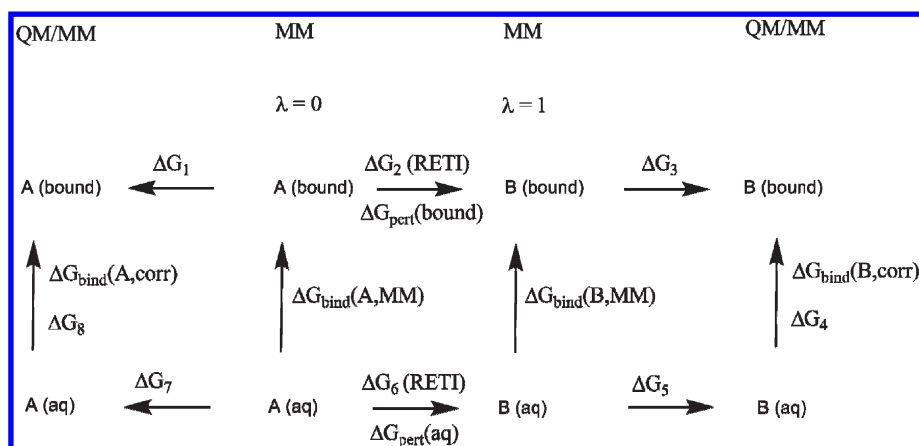
relative binding free energy of ligand “OH” compared to ligand “CH₃”, both bound to Cox-2, is calculated to be $\Delta \Delta G = -3.08 \pm 0.08$ kcal·mol⁻¹ for our standard sampling scheme (“repeated sampling”). Our alternative sampling scheme (“extended sampling”) yields a $\Delta \Delta G$ value of -2.87 ± 0.11 kcal·mol⁻¹. Both results are in good agreement with numbers previously reported for this perturbation: -2.99 ± 0.07^{13} and -2.9 ± 0.37^9 kcal·mol⁻¹.

The free energies with QM/MM corrections are summarized in Tables 2 (“repeated sampling”) and 3 (“extended sampling”). For repeated sampling, the QM/MM-corrected total binding free energy $\Delta G_4 - \Delta G_8$ of -4.5 ± 0.5 kcal·mol⁻¹ is in good agreement with the experimental relative binding free energy, which has been reported to be more negative than -4.64 kcal·mol⁻¹.⁹⁴ The corresponding number for extended sampling is -4.4 ± 0.6 kcal·mol⁻¹, a very similar value.

The sign of the free energies calculated according to eq 8 (ΔG_1 , ΔG_3 , ΔG_5 , and ΔG_7) indicates whether or not going from an MM representation without polarization to a QM/MM representation with polarization is energetically favorable. A negative free energy means that taking polarization into account is energetically favorable; i.e., the ligand was underpolarized in the classical force field. A positive free energy means the MM charges were such that the ligand was overpolarized in the MM model.

For our normal sampling scheme with completely independent runs (Table 2), the positive free energy ΔG_1 means that the MM charges for ligand OH in the bound situation are too large (ligand overpolarized) for the hydrophobic pocket in which the perturbation is performed. For ligand OH in water (“free”), the negative free energy (ΔG_7) indicates that the MM charges are too small to account for the real situation in water (ligand

Scheme 6. Thermodynamic Cycle with QM/MM Corrections Used for the Calculation of Relative Free Binding Energies of Protein–Ligand Complexes



underpolarized). Negative free energies ΔG_3 and ΔG_5 indicate that the MM charges for ligand CH_3 are equally underpolarized to reflect the situation both in the bound and the free state.

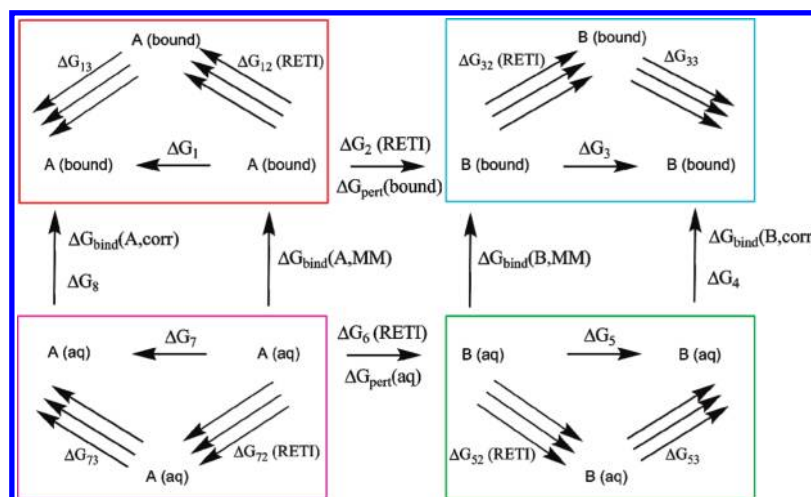
Our alternative sampling scheme (Table 3) also indicates that ligands OH and CH_3 are both underpolarized in the free state and that ligand OH is overpolarized in the hydrophobic pocket of the enzyme active site in the bound state. However, in contrast to “repeated sampling”, the results with “extended sampling” show that ligand CH_3 , too, is overpolarized in the hydrophobic pocket of the enzyme active site in the bound state.

The slightly different results for the alternative sampling schemes and the comparatively large standard errors for the QM/MM FEP free energies compared to the classical RETI energies indicate that it is more difficult to obtain converged results for the much more complex protein–ligand potential surfaces than for the simple test case methane/water. In particular, if the MM and the QM/MM potential surfaces are very different in shape, convergence problems may occur. It is also more difficult to ensure sufficient sampling in the case of the complex protein–ligand potential surfaces—according to the nature of the Zwanzig equation (exponential averaging) single configurations which may or may not be found can have a large impact on the final free energy. Histograms of the difference in solute–environment interaction between a QM/MM and an MM representation of the system ($\Delta U = U_{\text{QM/MM}} - U_{\text{QM,vac}} - U_{\text{charges,MM}} - U_{\text{Coul,solute-solv/protein,MM}}$ in eq 8) and histograms of the corresponding exponential terms are given in Figures S4–S7 in the Supporting Information. These histograms show that, owing to the form of the Zwanzig equation (eqs 2 and 8), a small percentage of configurations with a negative value of ΔU (the last column in Tables 2 and 3) can have a large effect on the final ΔG value as exponential averaging is performed. Finding such rare configurations is a challenge for the sampling method employed. Differences in the free energies obtained from QM/MM-FEP performed on the two ensembles obtained with different sampling schemes (ΔG_1 , ΔG_3 , ΔG_5 , ΔG_7 in Tables 2 and 3) can be explained by such sampling effects.

Interestingly, the mean values of ΔU given in Tables 2 and 3 are all positive; i.e., the QM/MM Hamiltonian employed in this study generally assesses the solute–environment interaction as less favorable than the MM Hamiltonian. Negative values of the corresponding free energy ΔG are nonetheless obtained because a relatively small percentage of configurations with negative ΔU

values dominate in eq 8. Whether the QM/MM Hamiltonian or the MM Hamiltonian gives a more “truthful” description of the reality and a further investigation of compatibility issues between MM force fields and QM Hamiltonians is beyond the scope of this study. Here we focus on investigating the numerical stability and convergence of a simple one-step QM/MM-FEP approach, which will be examined further in the next section. Future studies will certainly have to refine the approach and examine the compatibility of Hamiltonians.

Protein–Ligand System: Charge Perturbations. As shown above, generating alternative pathways by scaling solute charges up/down can be used to assess the pathway independence of the free energies calculated. Again, alternative configurations for the protein–ligand system were generated by performing RETI calculations in which solute charges were scaled up/down. In contrast to the charge perturbations for the methane/water system (in which no internal nonbonding interactions are evaluated as only 1,3-interactions are present), in the protein/ligand case, only the charges used for solute–environment interactions were scaled, while the charges used for ligand internal energy computation remained at their $\lambda = 0$ values. Eight λ windows (0.00, 0.14, 0.28, 0.42, 0.57, 0.72, 0.86, 1.00) were used to ensure a smooth transition between the Hamiltonians. The following charge scale factors were investigated: 0.80, 0.90, 0.95, 1.01, 1.05, 1.10, 1.20 (1.00: normal simulation without charge perturbation, see above). The last configurations from the normal, noncharge perturbation simulations presented above were taken as initial structures. Ten million equilibration moves were performed at each λ window (100 simulation blocks). For data collection, 30 million Monte Carlo moves were used (100 000 conformations \times 300 simulation blocks). One QM/MM single-point calculation was performed for each RETI simulation block, i.e., for 100 000 MM moves. All other simulation parameters were chosen analogously to the noncharge perturbation simulations described above. Equation 8 was used to calculate MM \rightarrow QM/MM FEP energies. The numbers thus obtained were used for comparison with the results obtained using the standard sampling scheme (“repeated sampling”) in the normal, noncharge perturbation RETI calculations. The thermodynamic cycle for the protein–ligand charge perturbations is depicted in Scheme 7.

Scheme 7. Thermodynamic Cycle with Charge Perturbations for the Protein–Ligand System^a

^a “A” represents ligand OH, “B” represents ligand CH₃. Red and cyan boxes: generation of alternative configurations by perturbing charges used for solute–environment interaction for the bound leg. Magenta and green boxes: generation of alternative configurations by perturbing charges used for solute–environment interaction for the free leg.

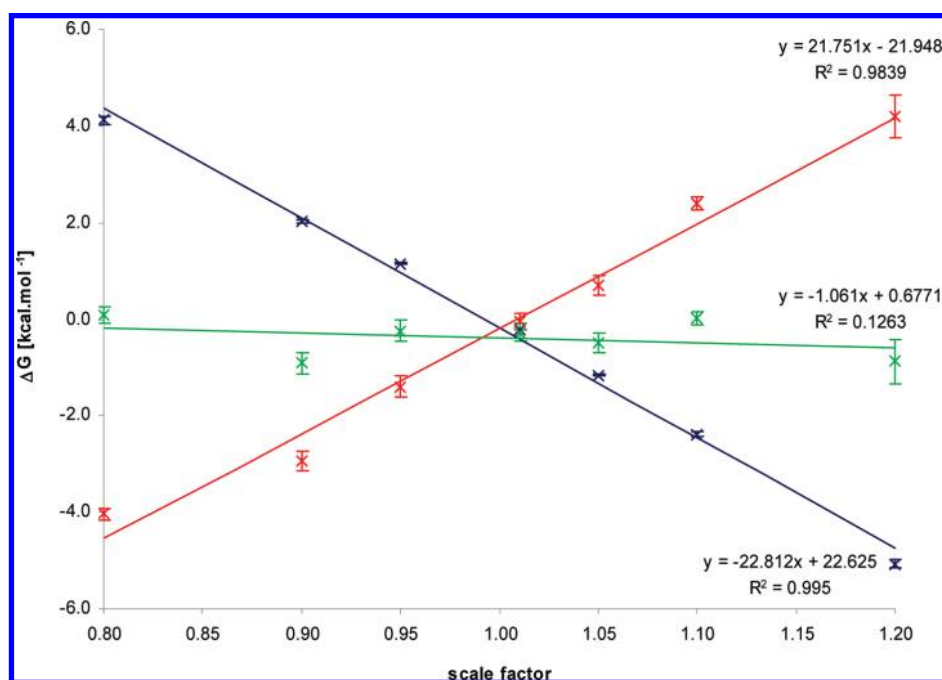


Figure 3. Free energies obtained from charge perturbations for ligand OH, bound to Cox-2. Blue: classical RETI energies (ΔG_{12}). Red: QM/MM-FEP energies calculated by eq 8 (ΔG_{13}). Green: sum of classical and QM/MM-FEP energies (thermodynamic cycle closure). Mean over all scale factors $\langle \Delta G_{12} + \Delta G_{13} \rangle = -0.39 \pm 0.14 \text{ kcal} \cdot \text{mol}^{-1}$; $\Delta G_1 = 0.28 \pm 0.23 \text{ kcal} \cdot \text{mol}^{-1}$.

The pathway independence of the free energy requires that for all pathways investigated

$$\Delta G_{72} + \Delta G_{73} = \Delta G_7 (\text{OH, free}) \quad (13)$$

$$\Delta G_{12} + \Delta G_{13} = \Delta G_1 (\text{OH, bound}) \quad (10)$$

$$\Delta G_{32} + \Delta G_{33} = \Delta G_3 (\text{CH}_3, \text{bound}) \quad (11)$$

$$\Delta G_{52} + \Delta G_{53} = \Delta G_5 (\text{CH}_3, \text{free}) \quad (12)$$

Cox-2 Charge Perturbations: Results. The free energies obtained from the charge perturbations of the Cox-2/inhibitor system are plotted in Figures 3–6.

In Figures 3–6, the sums of $\Delta G_{12} + \Delta G_{13}$, $\Delta G_{32} + \Delta G_{33}$, $\Delta G_{52} + \Delta G_{53}$, and $\Delta G_{72} + \Delta G_{73}$ are generally constant (small slopes of the green fitting lines). However, if the free energies are pathway independent, the mean values of these sums calculated over all scale factors must be equal to the corresponding noncharge

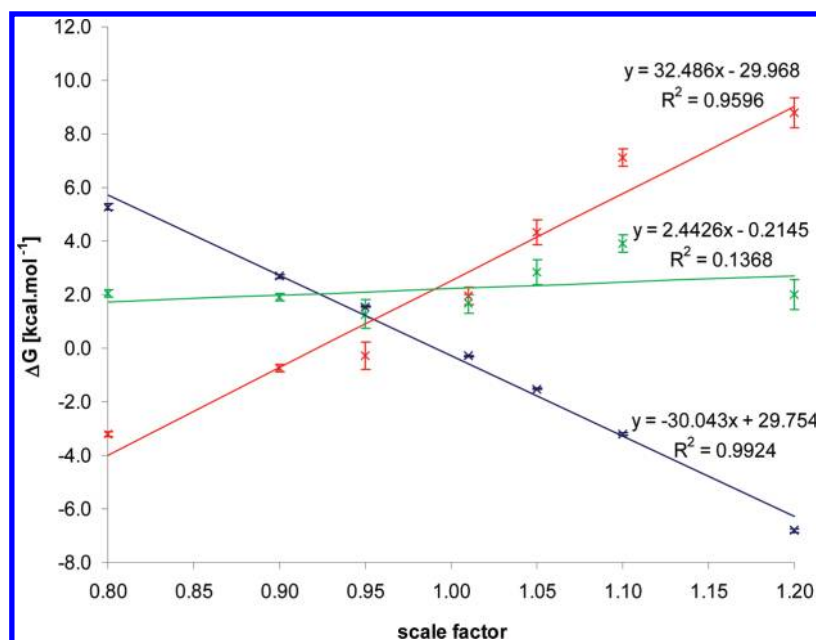


Figure 4. Free energies obtained from charge perturbations for ligand CH_3 , bound to Cox-2. Blue: classical RETI energies (ΔG_{32}). Red: QM/MM-FEP energies calculated by eq 8 (ΔG_{33}). Green: sum of classical and QM/MM-FEP energies (thermodynamic cycle closure). Mean over all scale factors $\langle \Delta G_{32} + \Delta G_{33} \rangle = 2.23 \pm 0.31 \text{ kcal} \cdot \text{mol}^{-1}$; $\Delta G_3 = -0.64 \pm 0.28 \text{ kcal} \cdot \text{mol}^{-1}$.

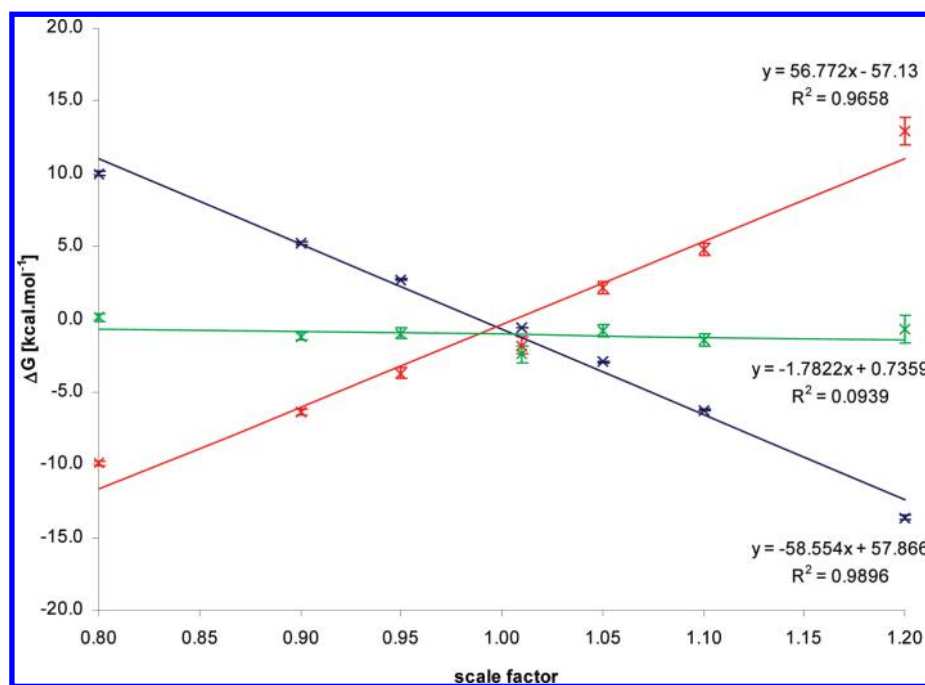


Figure 5. Free energies obtained from charge perturbations for ligand OH , solvated in water. Blue: classical RETI energies (ΔG_{72}). Red: QM/MM-FEP energies calculated by eq 8 (ΔG_{73}). Green: sum of classical and QM/MM-FEP energies (thermodynamic cycle closure). Mean over all scale factors $\langle \Delta G_{72} + \Delta G_{73} \rangle = -1.05 \pm 0.27 \text{ kcal} \cdot \text{mol}^{-1}$; $\Delta G_7 = -1.18 \pm 0.27 \text{ kcal} \cdot \text{mol}^{-1}$.

perturbation QM/MM free energies ΔG_1 , ΔG_3 , ΔG_5 , and ΔG_7 (eqs 10–13). The relevant free energies are summarized in Table S2 in the Supporting Information. We shall first discuss the results for the normal sampling scheme in the charge perturbations (one run) and “repeated sampling” in the noncharge perturbation simulations (10 independent runs). The above condition is perfectly fulfilled in the case of free ligand

OH , where $\langle \Delta G_{72} + \Delta G_{73} \rangle = -1.05 \pm 0.27 \text{ kcal} \cdot \text{mol}^{-1}$ and $\Delta G_7 = -1.18 \pm 0.27 \text{ kcal} \cdot \text{mol}^{-1}$. For free ligand CH_3 , $\langle \Delta G_{52} + \Delta G_{53} \rangle = -0.13 \pm 0.31 \text{ kcal} \cdot \text{mol}^{-1}$, which is still almost within the error bars of $\Delta G_5 = -0.71 \pm 0.26 \text{ kcal} \cdot \text{mol}^{-1}$. For bound ligand OH , $\langle \Delta G_{12} + \Delta G_{13} \rangle = -0.39 \pm 0.14 \text{ kcal} \cdot \text{mol}^{-1}$ is not in perfect agreement with $\Delta G_1 = 0.28 \pm 0.23 \text{ kcal} \cdot \text{mol}^{-1}$. The agreement is especially poor in the case of the bound CH_3

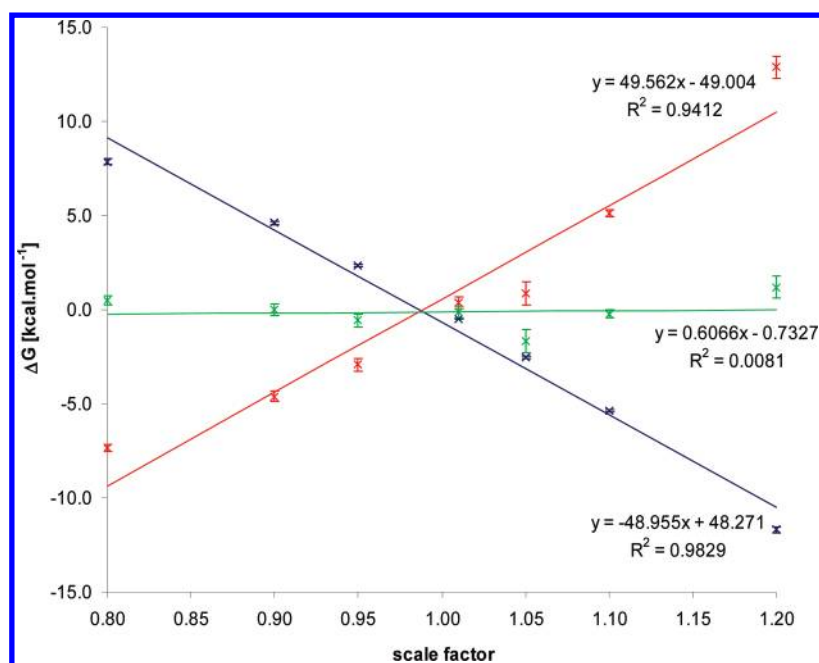


Figure 6. Free energies obtained from charge perturbations for ligand CH_3 , solvated in water. Blue: classical RETI energies (ΔG_{52}). Red: QM/MM-FEP energies calculated by eq 8 (ΔG_{53}). Green: sum of classical and QM/MM-FEP energies (thermodynamic cycle closure). Mean over all scale factors $\langle \Delta G_{52} + \Delta G_{53} \rangle = -0.13 \pm 0.31 \text{ kcal} \cdot \text{mol}^{-1}$; $\Delta G_5 = -0.71 \pm 0.26 \text{ kcal} \cdot \text{mol}^{-1}$.

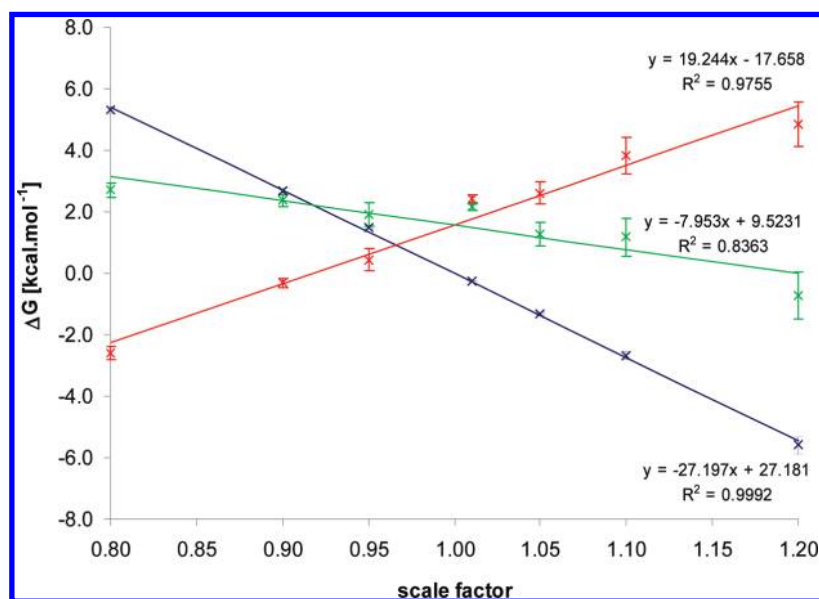


Figure 7. Free energies obtained from charge perturbations for ligand CH_3 , bound to Cox-2, extended sampling scheme (five runs, using the last restart file from the previous run as initial configuration). Blue: classical RETI energies (ΔG_{32}). Red: QM/MM-FEP energies calculated by eq 8 (ΔG_{33}). Green: sum of classical and QM/MM-FEP energies (thermodynamic cycle closure). Mean over all scale factors $\langle \Delta G_{32} + \Delta G_{33} \rangle = 1.56 \pm 0.40 \text{ kcal} \cdot \text{mol}^{-1}$; ΔG_3 (extended sampling) = $1.02 \pm 0.41 \text{ kcal} \cdot \text{mol}^{-1}$.

ligand, where $\langle \Delta G_{32} + \Delta G_{33} \rangle = 2.23 \pm 0.31 \text{ kcal} \cdot \text{mol}^{-1}$ and $\Delta G_3 = -0.64 \pm 0.28 \text{ kcal} \cdot \text{mol}^{-1}$.

These results indicate that it is much more difficult to obtain converged results for a ligand bound to a protein than in aqueous solution, probably due to a more frustrated potential energy surface for a protein–ligand complex. Also, ligand CH_3 seems to be more challenging than ligand OH , although this may very well purely reflect chance events in the simulation sampling.

Additional Charge Perturbations for Ligand CH_3 ("Extended Sampling"). As the charge perturbations of ligand CH_3 revealed the need to improve sampling (see above) and to provide data for comparison with the extended sampling scheme employed in the normal noncharge perturbation simulations, we additionally developed an extended sampling scheme for the charge perturbations of ligand CH_3 : for this ligand only, charge perturbations were repeated five times (300 blocks each), using the last restart file from the previous run as the initial configuration.

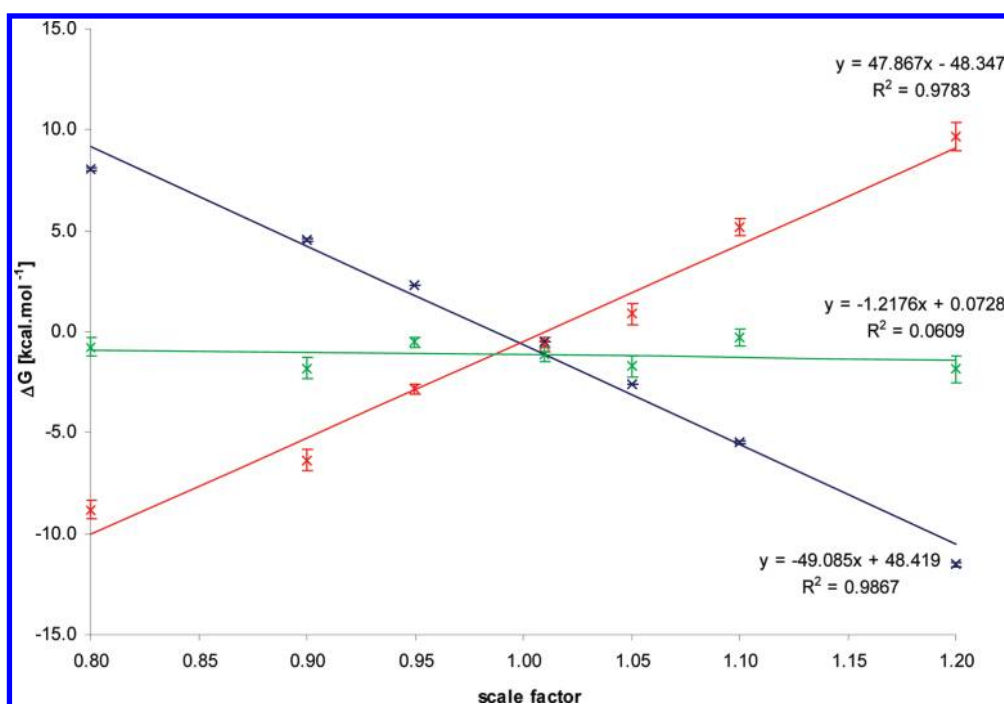


Figure 8. Free energies obtained from charge perturbations for ligand CH_3 , solvated in water, extended sampling scheme (five runs, using the last restart file from the previous run as initial configuration). Blue: classical RETI energies (ΔG_{52}). Red: QM/MM-FEP energies calculated by eq 8 (ΔG_{53}). Green: sum of classical and QM/MM-FEP energies (thermodynamic cycle closure). Mean over all scale factors $\langle \Delta G_{52} + \Delta G_{53} \rangle = -1.15 \pm 0.23 \text{ kcal} \cdot \text{mol}^{-1}$; ΔG_5 (extended sampling) $= -1.04 \pm 0.28 \text{ kcal} \cdot \text{mol}^{-1}$.

In this case, the RETI energies reported are the mean values of five runs; FEP energies are calculated using eq 8 on all 1500 configurations.

The results are shown in Figures 7 and 8. While the gradients of the fitting lines for the sums $\Delta G_{32} + \Delta G_{33}$ and $\Delta G_{52} + \Delta G_{53}$ are not ideal (i.e., zero), the mean values of the sums over all scale factors show a much better agreement with the corresponding noncharge perturbation free energies for extended sampling. In the case of bound ligand CH_3 , the mean over all scale factors $\langle \Delta G_{32} + \Delta G_{33} \rangle = 1.56 \pm 0.40 \text{ kcal} \cdot \text{mol}^{-1}$ while ΔG_3 (extended sampling) $= 1.02 \pm 0.41 \text{ kcal} \cdot \text{mol}^{-1}$. For free ligand CH_3 , the mean over all scale factors $\langle \Delta G_{52} + \Delta G_{53} \rangle = -1.15 \pm 0.23 \text{ kcal} \cdot \text{mol}^{-1}$ and ΔG_5 (extended sampling) $= -1.04 \pm 0.28 \text{ kcal} \cdot \text{mol}^{-1}$. These numbers show that, using the extended sampling scheme, the free energies obtained via alternative pathways (RETI charge perturbations plus QM/MM-FEP) are within the error bars of the free energies obtained from a “normal” (i.e., without charge perturbations) QM/MM-FEP calculation, although the nonzero gradient of the sums of $\Delta G_{32} + \Delta G_{33}$ (green line in Figure 7) shows that in the particular case of bound ligand CH_3 even more sampling would be desirable for perfect thermodynamic cycle closure. If the furthest outlier, the data point at scale factor 1.20, is omitted, a smaller gradient for $\Delta G_{32} + \Delta G_{33}$ is observed ($y = -5.18x + 6.96$, plot not shown). We emphasize that the interval of charge scale factors used in this study was chosen completely arbitrarily; no prediction is possible about reasonable values for these factors, and it seems likely that the further one deviates from the normal MM parameters at a scale factor of 1.0, the more likely the Zwanzig equation will show poorer convergence. Furthermore, the strong dependence of convergence on the sampling scheme (“repeated” vs “extended” sampling) chosen further emphasizes the need to find rare configurations by extensively sampling the complex protein–ligand potential surface—a priori there is no reason to prefer one sampling scheme to another.

Our results indicate that doing FEP in one step, i.e., without intermediate states, between an MM and a QM/MM representation of the system, is a reasonable approximation at moderate computational cost, even in the most challenging situation where the solute and parts of the protein environment are treated flexibly, and that only sampling the MM potential surface is justifiable, at least in the case of the systems studied in this work. However, convergence is only achieved if extensive sampling of the partially flexible protein–ligand system is performed.

CONCLUSIONS AND OUTLOOK

The aim of the current study was not the exact reproduction of experimental solvation/binding free energies, but to investigate whether free energy perturbation between a QM/MM and an MM representation of a system can be performed in one step, i.e., without intermediate states. A simple implementation was chosen deliberately, using existing codes for MM and QM/MM and coupling these by a simple scripting language.

For the water/methane system, the QM/MM-corrected relative free energy of hydration, while not in particularly good agreement with experiment, is precise and reproducible. For the Cox-2 protein–ligand test system, applying our approach yields a QM/MM-corrected relative binding free energy in closer agreement with the experimentally determined value, which is probably fortuitous. Applying the method to a wider range of protein–ligand systems is required to achieve further insights. We observe that treating a solute as a rigid entity leads to well-converged results, an observation which has been made previously by other authors.^{40,41,53,55} It is a much greater challenge to achieve converged free energy changes with a flexible protein–ligand complex, and extensive sampling is required in this case.

The investigation of the distribution of the difference of the solute–environment interaction energies given by the two Hamiltonians used and the corresponding histograms of the exponential terms in the Zwanzig equation shows that it is essential to detect rare configurations in the sampling process; this emphasizes the need to extensively sample the potential surface of interest.

The free energy calculations performed here rely on the use of the Zwanzig equation in one direction only. Ordinarily, to demonstrate convergence, the reverse perturbations (in this case from a QM/MM ensemble to the MM Hamiltonian) would be performed, with the expectation that the two results should be equal and opposite. The methodology presented here does not allow this test. However, given that free energy is a state function, and that the calculated free energy change from the MM to the QM/MM Hamiltonian should therefore be pathway independent, we have modified the MM ensemble by scaling the electrostatic parameters and demonstrated that indeed the overall calculated free energy change is consistent and unaffected. While this is not a sufficient condition to demonstrate the appropriate use of the Zwanzig equation in one step, it is a necessary one, and the best we can offer. Thus, in terms of providing recommendations for the use of this approach, demonstrating pathway independence for the calculated free energies, by scaling the underlying MM parameters, is a requirement. For other systems where we have used our simple approach, pathway independence has been shown. In addition, the free energies calculated by our approximate method, and independently by the more rigorous method of Woods et al.,³⁸ are identical to within statistical error.⁹⁸

Crude approximations were made in our current model and should be kept in mind: in the QM/MM part of the protocol, the MM environment was represented by point charges that polarize the QM part, no back-polarization was considered, and a cutoff was used both for the MM part of the simulations and for QM/MM. Only a small fraction of MM configurations was chosen for post processing with QM/MM (every 100 000th). Sampling was only performed for the MM ensemble; unlike other approaches, no QM/MM sampling was performed. This approximation has the advantage that a simple postprocessing of MM configurations by a QM/MM code can be performed, without the need of creating a computationally expensive QM/MM ensemble.

In principle, the QM/MM-FEP free energies can be directly interpreted: they show if an MM ligand is overpolarized or underpolarized in a given binding/solvation context. Overall, our approach allows electronic polarization effects to be estimated in a simple way. However, more research is required for a reliable and robust application to diverse protein–ligand systems. The optimum combination of an MM force field with a QM level of theory needs to be determined, and some reparametrization of the MM model may well be required. This work is currently in progress.

■ ASSOCIATED CONTENT

S Supporting Information. Details of the following aspects of this work: (1) brief literature review; (2) plot of MM gradients for the charge perturbations of water and methane; (3) plot showing free energies obtained from charge perturbations for water (B3LYP/6-31G*); (4) plot showing free energies obtained from charge perturbations for methane (B3LYP/6-31G*); (5) table with free energies and standard errors as defined in Scheme 3 for

the system methane-water; (6) detailed protocol for the MM free energy calculations (protein–ligand system); (7) table with free energies and standard errors for different pathways as defined in Scheme 7 for the perturbation of ligand OH to CH₃, solvated in water (“free”), and bound to Cox-2 (“bound”); (8) histograms of the difference of solute-protein/solvent interaction energies in the QM/MM case and in the MM case, as defined in eq 8 for the sampling scheme “repeated sampling”; (9) histograms of the difference of solute–protein/solvent interaction energies in the QM/MM case and in the MM case, as defined in eq 8 for the sampling scheme “extended sampling”; (10) histograms of the exponential term in eq 8 for the sampling scheme “repeated sampling”; and (11) histograms of the exponential term in eq 8 for the sampling scheme “extended sampling” (16 pages). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +44 238059 2794. Fax: +44 238059 3781. E-mail: j.w.essex@soton.ac.uk.

Present Addresses

[†](a) Computer-Chemie-Centrum, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstrasse 25, 91052 Erlangen, Germany, and (b) Excellence Cluster “Engineering of Advanced Materials”, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstrasse 49b, 91052 Erlangen, Germany.

[‡]School of Chemistry, University of Edinburgh, Joseph Black Building, The King’s Buildings, West Mains Road, Edinburgh EH9 3JJ, UK.

■ ACKNOWLEDGMENT

This work was supported by the BBSRC as part of the IntBioSim project. The authors thank Prof. Adrian Mulholland, Dr. Christopher Woods, Dr. Brian Cheney, and Mag. Pharm. Patrick Schopf for many helpful discussions and advice.

■ REFERENCES

- (1) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.
- (2) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 15–26.
- (3) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151–166.
- (4) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. *J. Med. Chem.* **2006**, *49*, 5851–5855.
- (5) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J. Med. Chem.* **2006**, *49*, 5912–5931.
- (6) Gilson, M. K.; Zhou, H.-X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
- (7) Michel, J.; Essex, J. W. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 639–658.
- (8) Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A. *Nature* **1987**, *328*, 551–554.
- (9) Michel, J.; Verdonk, M. L.; Essex, J. W. *J. Med. Chem.* **2006**, *49*, 7427–7439.
- (10) Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; Anderson, K. S.; Hamilton, A. D. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 663–667.

- (11) Kim, J. T.; Hamilton, A. D.; Bailey, C. M.; Domoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2006**, *128*, 15372–15373.
- (12) Oostenbrink, C.; van Gunsteren, W. F. *Proteins: Struct., Funct., Bioinf.* **2004**, *54*, 237–246.
- (13) Michel, J.; Verdonk, M. L.; Essex, J. W. *J. Chem. Theory Comput.* **2007**, *3*, 1645–1655.
- (14) Michel, J.; Essex, J. W. *J. Med. Chem.* **2008**, *51*, 6654–6664.
- (15) Michel, J.; Taylor, R. D.; Essex, J. W. *J. Chem. Theory Comput.* **2006**, *2*, 732–739.
- (16) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13703–13710.
- (17) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13711–13718.
- (18) Mackerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
- (19) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (20) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (21) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (22) Lee, T.-S.; Lewis, J. P.; Yang, W. *Comput. Mater. Sci.* **1998**, *12*, 259–277.
- (23) van der Vaart, A.; Gogonea, V.; Dixon, S. L.; Merz, K. M., Jr. *J. Comput. Chem.* **2000**, *21*, 1494–1504.
- (24) van der Vaart, A.; Suarez, D.; Merz, K. M., Jr. *J. Chem. Phys.* **2000**, *113*, 10512–10523.
- (25) Dixon, S. L.; Merz, K. M., Jr. *J. Chem. Phys.* **1996**, *104*, 6643–6649.
- (26) Dixon, S. L.; Merz, K. M., Jr. *J. Chem. Phys.* **1997**, *107*, 879–893.
- (27) Stewart, J. J. P. *J. Mol. Model.* **2009**, *15*, 765–805.
- (28) Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133–146.
- (29) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085–1123.
- (30) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (31) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523–2544.
- (32) Monard, G.; Merz, K. M. *Acc. Chem. Res.* **1999**, *32*, 904–911.
- (33) Bathelt, C. M.; Zurek, J.; Mulholland, A. J.; Harvey, J. N. *J. Am. Chem. Soc.* **2005**, *127*, 12900–12908.
- (34) Ridder, L.; Harvey, J. N.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. *J. Phys. Chem. B* **2003**, *107*, 2118–2126.
- (35) Beierlein, F.; Lanig, H.; Schürer, G.; Horn, A. H. C.; Clark, T. *Mol. Phys.* **2003**, *101*, 2469–2480.
- (36) Beierlein, F. R.; Othersen, O. G.; Lanig, H.; Schneider, S.; Clark, T. *J. Am. Chem. Soc.* **2006**, *128*, 5142–5152.
- (37) Beierlein, F.; Clark, T. In *High Performance Computing in Science and Engineering, Munich 2004—Transactions of the Second Joint HLRB and KONWIHR Status and Result Workshop, March 2–3, 2004, Technical University of Munich and Leibnitz-Rechenzentrum Munich*; Wagner, S., Hanke, W., Bode, A., Durst, F., Eds.; Springer Verlag: Berlin, 2004; pp 245–260.
- (38) Woods, C. J.; Manby, F. R.; Mulholland, A. J. *J. Chem. Phys.* **2008**, *128*, 014109.
- (39) Rod, T. H.; Rydberg, P.; Ryde, U. *J. Chem. Phys.* **2006**, *124*, 174503.
- (40) Rod, T. H.; Ryde, U. *Phys. Rev. Lett.* **2005**, *94*, 138302.
- (41) Rod, T. H.; Ryde, U. *J. Chem. Theory Comput.* **2005**, *1*, 1240–1251.
- (42) Liu, W.; Sakane, S.; Wood, R. H.; Doren, D. J. *J. Phys. Chem. A* **2002**, *106*, 1409–1418.
- (43) Ming, Y.; Lai, G.; Tong, C.; Wood, R. H.; Doren, D. J. *J. Chem. Phys.* **2004**, *121*, 773–777.
- (44) Sakane, S.; Liu, W.; Doren, D. J.; Shock, E. L.; Wood, R. H. *Geochim. Cosmochim. Acta* **2001**, *65*, 4067–4075.
- (45) Sakane, S.; Yezdimer, E. M.; Liu, W.; Barriocanal, J. A.; Doren, D. J.; Wood, R. H. *J. Chem. Phys.* **2000**, *113*, 2583–2593.
- (46) Wood, R. H.; Liu, W.; Doren, D. J. *J. Phys. Chem. A* **2002**, *106*, 6689–6693.
- (47) Wood, R. H.; Yezdimer, E. M.; Sakane, S.; Barriocanal, J. A.; Doren, D. J. *J. Chem. Phys.* **1999**, *110*, 1329–1337.
- (48) Bandyopadhyay, P. J. *J. Chem. Phys.* **2005**, *122*, 091102.
- (49) Iftimie, R.; Salahub, D.; Schofield, J. *J. Chem. Phys.* **2003**, *119*, 11285–11297.
- (50) Iftimie, R.; Salahub, D.; Wei, D.; Schofield, J. *J. Chem. Phys.* **2000**, *113*, 4852–4862.
- (51) Iftimie, R.; Schofield, J. *J. Chem. Phys.* **2001**, *114*, 6763–6773.
- (52) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *ChemPhysChem* **2009**, *10*, 112–1134.
- (53) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009**, *113*, 1253–1272.
- (54) Rosta, E.; Haranczyk, M.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **2008**, *112*, 5680–5692.
- (55) Rosta, E.; Klähn, M.; Warshel, A. *J. Phys. Chem. B* **2006**, *110*, 2934–2941.
- (56) Olsson, M. H. M.; Hong, G.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025–5039.
- (57) Štrajbl, M.; Hong, G.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 13333–13343.
- (58) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293–2301.
- (59) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516–17524.
- (60) Wesolowski, T.; Warshel, A. *J. Phys. Chem.* **1994**, *98*, 5183–5187.
- (61) Vaidehi, N.; Wesolowski, T. A.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 4264–4271.
- (62) Heimdal, J.; Rydberg, P.; Ryde, U. *J. Phys. Chem. B* **2008**, *112*, 2501–2510.
- (63) Kaukonen, M.; Söderhjelm, P.; Heimdal, J.; Ryde, U. *J. Chem. Theory Comput.* **2008**, *4*, 985–1001.
- (64) Shaw, K. E.; Woods, C. J.; Mulholland, A. J. *J. Phys. Chem. Lett.* **2009**, *1*, 219–223.
- (65) Hastings, W. K. *Biometrika* **1970**, *57*, 97–109.
- (66) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (67) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146.
- (68) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
- (69) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, 2006.
- (70) Woods, C.; Michel, J. *ProtoMS 2.2*; University of Southampton: Southampton, UK, 2002–2007.
- (71) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117–9133.
- (72) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (73) Perl website: <http://www.perl.org/>.
- (74) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.
- (75) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (76) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724–728.

- (77) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (78) Hariharan, P. C.; Pople, J. A. *Theor. Chem. Acc.* **1973**, *28*, 213–222.
- (79) Hariharan, P. C.; Pople, J. A. *Mol. Phys.* **1974**, *27*, 209–214.
- (80) Gordon, M. S. *Chem. Phys. Lett.* **1980**, *76*, 163–168.
- (81) Francel, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; DeFrees, D. J.; Pople, J. A.; Gordon, M. S. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
- (82) Binning, R. C., Jr.; Curtiss, L. A. *J. Comput. Chem.* **1990**, *11*, 1206–1216.
- (83) Blaudeau, J.-P.; McGrath, M. P.; Curtiss, L. A.; Radom, L. *J. Chem. Phys.* **1997**, *107*, 5016–5021.
- (84) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. *J. Chem. Phys.* **1998**, *109*, 1223–1229.
- (85) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. *J. Comput. Chem.* **2001**, *22*, 976–984.
- (86) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (87) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (88) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (89) Peterson, K. A.; Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 7410–7415.
- (90) Davidson, E. R. *Chem. Phys. Lett.* **1996**, *260*, 514–518.
- (91) Alavi, S.; Ripmeester, J. A.; Klug, D. D. *J. Chem. Phys.* **2007**, *126*, 124708.
- (92) El-Sheikh, S. M.; Barakat, K.; Salem, N. M. *J. Chem. Phys.* **2006**, *124*, 124517.
- (93) Plouunt Price, M. L.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 9455–9466.
- (94) Penning, T. D.; Talley, J. J.; Bertenshaw, S. R.; Carter, J. S.; Collins, P. W.; Docter, S.; Graneto, M. J.; Lee, L. F.; Malecha, J. W.; Miyashiro, J. M.; Rogers, R. S.; Rogier, D. J.; Yu, S. S.; Anderson, G. D.; Burton, E. G.; Cogburn, J. N.; Gregory, S. A.; Koboldt, C. M.; Perkins, W. E.; Seibert, K.; Veenhuizen, A. W.; Zhang, Y. Y.; Isakson, P. C. *J. Med. Chem.* **1997**, *40*, 1347–1365.
- (95) Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. *J. Comput. Chem.* **1993**, *14*, 1240–1249.
- (96) Udier-Blagovic, M.; Tirado, P. M. D.; Pearlman, S. A.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 1322–1332.
- (97) Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **1997**, *101*, 9663–9669.
- (98) Carter, M. K.; Schopf, P.; Essex, J. W. Unpublished results.