Implicit versus explicit solvent in free energy calculations of enzyme catalysis: Methyl transfer catalyzed by catechol O-methyltransferase

Rod, Thomas; Rydberg, Patrik; Ryde, Ulf

Published in:
Journal of Chemical Physics

DOI:
10.1063/1.2186635

2006

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

Implicit versus explicit solvent in free energy calculations of enzyme catalysis: methyl transfer catalyzed by catechol O-methyltransferase

Thomas H. Rod[*,‡], Patrik Rydberg, and Ulf Ryde

Department of Theoretical Chemistry, Chemical Center, Lund University, P.O. Box 124, S-22100 Lund, Sweden.

4/10/17

Abstract

We compare free energy calculations for the methyl transfer reaction catalyzed by catechol O-methyltransferase models using the quantum mechanical/molecular mechanical free energy (QM/MM-FE) method with implicit and explicit solvent. An analogous methylation reaction in solution is also studied. For the explicit solvent model, we use the transferable intermolecular potential three-point (TIP3P) model, and for the implicit model, we use the molecular volume generalized Born model as implemented in CHARMM. We find that activation and reaction free energies calculated with the two models are very similar, despite that some structural differences exist. A significant change in the polarization of the environment occurs as the reaction proceeds, which is more pronounced for the reaction in solution than for the enzymatic reaction. For the

[*] Corresponding author: E-mail: tr@atomistix.com. Phone: +45 35 320 630.

[‡] Current address: Atomistix A/S, Julianne Maries Vej 30, DK-2100 Copenhagen Ø, Denmark.

enzymatic reaction, most of the change takes place in the protein rather than the solvent, and, hence, the benefit of having instantaneous relaxation of the solvent degrees of freedom is less pronounced for the enzymatic reaction than for the reaction in solution. This is a likely reason why energies of the enzyme reaction are less sensitive to the choice of Born radii than are energies of the reaction in solution.

Introduction

Solvent has a crucial effect on many biochemical processes, and both structure and activity of enzymes should be understood in the context of the solvent. Without a solvent, many enzymes would be unstable in their active form. Therefore, it is important to properly model the solvent in theoretical studies of biochemical processes such as protein folding and enzyme catalysis. In atomic-scale simulations, explicit modeling of solvent molecules causes several complications. Not only does the number of atoms increase significantly, but the extra degrees of freedom also make it more difficult to obtain converged thermodynamic properties from simulations.

Implicit solvent models encompass an alternative way to model the effect of the solvent. The basic idea is to model the solute as a region with a low dielectric constant embedded in a solvent with a different and typically higher dielectric constant (about 80 for water). In this model, the solvation energy is decomposed into an electrostatic term and a non-polar term. Hence, the solvation energy reads:

$$\Delta G_{\text{solv}} = \Delta G_{\text{elec}} + \Delta G_{\text{np}} \quad (1)$$

The electrostatic component contains electrostatic interactions between solute-solvent charges as well as the energy required to polarize the solvent. The term can be computed by solving the Poisson equation:

$$\nabla[\epsilon(r)\nabla\phi(r)] = -4\pi\rho(r) \quad (2)$$

where $\rho(r)$ is the charge density, which is zero outside the solute, and $\phi(r)$ is the corresponding electrostatic potential. $\epsilon(r)$ is the dielectric constant, which has one constant value inside the solute and another constant value outside the solute. The

Poisson equation is only valid for zero salt concentration, but it can be extended to the more general Poisson-Boltzmann equation, which also covers nonzero salt concentrations.

The non-polar solvation term comes from the the combined effect of two types of interactions: the unfavorable energy required to create a cavity in the solvent with the same shape as the solute, and the favorable van der Waals interaction between the solute and solvent molecules. The solvent-accessible-surface approximation (SASA) is often used to compute this term. In this approximation, the non-polar term is computed by:

$$\Delta G_{\text{np}} = \sum_i \sigma_i A_i \quad (3)$$

where σ_i is an empirical parameter and A_i is the solvent accessible surface area of atom i.

Solving Poisson's equation provides the solvation free energy for a single configuration of the solute. The equation can be solved numerically and today many efficient algorithms and programs are developed for that purpose,[1] but despite these efforts, it is still too costly to solve the Poisson equation for many configurations as would be necessary for efficient sampling of the solute degrees of freedom. For these reasons, schemes based on solving the Poisson equation typically consider only a single or few configurations of the solute.

A more approximate, but also cheaper approach is to use the generalized Born (GB) model. In the GB model, the free energy obtained from solving the Poisson equation is approximated by a sum over pseudo pair potentials:

$$\Delta G_{\text{elec}} = \frac{1}{2}\left(1-\frac{1}{\epsilon}\right)\sum_{i,j} \frac{q_i q_j}{\sqrt{r_{i,j}^2 + \alpha_i \alpha_j \exp\left(-r_{i,j}/F\,\alpha_i \alpha_j\right)}} \quad (4)$$

where ϵ is the dielectric constant of the solvent, q_i and q_j are the charges of atom i and j, and $r_{i,j}$ is the distance between the two atoms. F is an empirical constant, originally proposed to be 4.[2] The individual terms in the above equation are not true pair potentials, since the so-called Born radii α_i and α_j depend on the shape of the solute, and therefore on the positions of all other atoms encompassing the solute. The Born radius of an atom can be computed exactly by solving the Poisson equation with the charges of all atoms set to zero except the one under consideration. In that case, it has been shown that the GB equation gives results in close agreement with those obtained from solving the Poisson equation.[3] Although this validates the GB equation, it is of little practical usage, since it demands solving the Poisson equation once for each atom of the solute in order to employ an approximation to a single Poisson equation. Instead, more approximate schemes are used to obtain the Born radii and the success of GB depends on the efficiency of these schemes to obtain accurate Born radii. Much effort has been put into this, resulting in several different implementations of GB with varying degrees of performance in terms of speed and accuracy.[1,4,5,6,7,8]

Relatively few studies have compared implicit solvent calculations with corresponding ones using an explicit solvent model and these have mostly considered protein folding. The comparative studies have been somewhat discouraging since the GB simulations often predict the wrong native structure compared to simulations in explicit solvent, as well as to experiment. In contrast, the predictions based on simulations in explicit solvent generally agree well with experiments.[7,8,9,10,11] It is possible that different and more accurate GB models would perform better than the ones employed in those studies. Indeed, a general problem in the above GB simulations seems to be either the lack of or

unwanted formation of salt bridges. Case and coworkers[6] found that a salt bridge, which was formed in the explicit solvent simulations, was not formed in simulations based on the GB/ACE variant,[12] whereas it was formed in simulations based on the GB/OGB variant developed by Onufriev, Bashford, and Case.[13]

Perhaps the greatest benefit of using implicit solvent is that the solvent relaxes instantaneously to a modification of the solute. This can be utilized with advantage in pKa calculations, in particular those where the response of the solute to a proton can be neglected. In the simplest approximation, pKa calculations can then be performed on a single structure of either the protonated or deprotonated complex. At a next level of approximation, it can be used in free energy perturbation calculations.[14,15] The immediate relaxation of the solvent is also exploited in the so-called MM/PBSA method,[16] and in constant pH simulations.[17,18,19,20]

Besides the simulations of protein folding, there are relatively few studies that compare implicit and explicit solvent simulations. One such study is the calculation of pKa values for different Asp residues in thioredoxin and ribonuclease[6]. In that study, it was found that GB/OBC gave results in better agreement with experiment than calculations based on explicit solvent simulations.

We are specifically interested in free energy calculations related to enzyme catalysis. In this article, we investigate the possibility of using GB to model solvent in a methyl-transfer reaction catalyzed by catechol O-methyltransferase (COMT) studied by the quantum mechanical/molecular mechanical free energy method (QM/MM-FE) used previously by Yang and coworkers,[15,21,22] Ishida and Kato,[23,24] as well as by us.[25] The COMT reaction has previously been used in our laboratory to compare various free-

energy methods.[25,26,27] We have shown that the QM/MM-FE method give results in close agreement with the more elaborate quantum-mechanical thermodynamic cycle perturbation (QTCP) method.[25] In the QM/MM-FE method, a reaction pathway is obtained by means of QM/MM optimization. In a subsequent step, simulations are performed for points along the reaction pathway, in which the QM region is fixed and is represented by point charges obtained from QM/MM calculations. These simulations are then used to compute the MM-QM interaction free energy, which is added to the QM energy obtained from the QM/MM calculations.[21,25] Since the MM-QM interaction energy is computed by simple Coulomb interactions between two sets of point charges, it is straightforward to use implicit solvent models in these calculations instead of solvent described explicitly.

Here, we make a comparative study of the generalized Born model, represented by the molecular volume generalized Born (GBMV) model[28,29] and the three-point transferable intermolecular potential (TIP3P) explicit solvent model.[30] The GBMV model has been found to reproduce solutions to the Poisson equation most accurately among different GB implementations and at the same time being faster than direct solvers of the Poisson equation.[1] Hence, GBMV is an excellent model for testing the ability of implicit solvent models to reproduce results based on explicit solvent, which currently is considered to be the most accurate solvent description.

GBMV differs from most GB methods in a number of ways. The Born radii, α_i , are computed by:

$$\frac{1}{\alpha_i}=\frac{1}{R_i}-\frac{1}{4\pi}\int\frac{1}{r^4}\mathrm{d}V+\left(\frac{1}{4R_i^4}-\frac{1}{4\pi}\int\frac{1}{r^7}\mathrm{d}V\right)^{1/4} \quad (5)$$

where R_i is the atomic radius of atom i. The above equation differs from the more commonly used Coulomb approximation, which corresponds to the first two terms on the right hand side in the above equation. The GBMV method also incorporates the non-polar solvation term (3) based on an approximation to the molecular surface.[29]

We compare free energy barriers for the reaction catalyzed by COMT as well as for a corresponding methyl transfer reaction in solution. We find that structural properties, such as hydrogen bonds, are somewhat different in the explicit and implicit solvent simulations. Despite these problems, the GBMV model reproduces the free energy barrier based on explicit solvent models very well, and the free energy calculations are robust towards changes in the atomic radii for the enzyme-catalyzed reaction. In contrast, the free energy barrier for the reaction in solution is more sensitive to the choice of the Born radii.

Method

Model system

Catechol O-methyltransferase (COMT) catalyzes the transfer of a methyl group from S-adenosylmethionine to the oxygen of a catechol group. The natural substrates of COMT are catecholamine neurotransmitters, like dopamine and (nor)adrenaline, which are deactivated by the methyl transfer. The crystal structure of the enzyme shows that the substrate binds directly to a Mg2+ ion with both its catechol oxygen atoms.[31] We have modeled the active site by a Mg2+ ion, a catecholate group with a single negative charge, a water, one formamide, and two formate groups, as models of the three Mg ligands from the protein, Asn-170, Asp-141, and Asp-169. In addition, S(CH3)3+ was used as a model of S-adenosylmethionine. This quantum system is illustrated in Figure 1a. The

surrounding enzyme was modeled by the Amber-94 force field, with coordinates taken from the crystal structure.[31] The five junctions between the QM and MM systems were treated by the standard hydrogen link-atom method.[32] In the MM calculations, it was assumed that all Asp and Glu residues are negatively charged, and that the Lys and Arg residues are positively charged. Based on the hydrogen-bond network around the two His residues, we decided to let His-139 to be protonated on the Ne2 atom, whereas His-190 was protonated on both nitrogen atoms and therefore was positively charged.
In the corresponding reaction in solution, we used the same reaction path, but only the catecholate and S(CH3)3+ molecules were considered (Figure 1b).

QM/MM-FE method

In the QM/MM-FE method, free energy changes are calculated between points along a reaction pathway, which is spanned by atoms in the QM region. A free energy change is computed as the sum of a QM energy change and a change in an interaction free energy between the MM and QM regions. The electrostatic contribution to the interaction free energy is approximated with interactions between point charges. Hence, in practice a QM/MM-FE calculation consists of a calculation of a reaction pathway by means of QM/MM. These calculations also result in a set of point charges of the QM region for each point along the reaction pathway, as well as an energy. In a subsequent step, an MD simulation is performed, in which the atoms in the QM region are fixed and described by the computed point charges. These simulations are finally used to compute MM-QM interaction free energy changes in steps along the reaction pathway by means of free energy perturbation. Thus,

$$\Delta A(i \rightarrow j) = \left\{ E_{\mathrm{QM}}(R_{\mathrm{QM}}^{(j)}) - E_{\mathrm{QM}}(R_{\mathrm{QM}}^{(i)}) \right\} + k_{\mathrm{B}} T \ln \left\langle \exp\left[-\left(U(R_{\mathrm{QM}}^{(j)}) - U(R_{\mathrm{QM}}^{(i)}) \right) / k_{\mathrm{B}} T \right] \right\rangle_{R_{\mathrm{QM}}^{(i)}} \quad (6)$$

In the QM/MM-FE approach, the QM energy terms enclosed by curled brackets are the energy of the QM region only, but polarized by the MM atoms. Hence, this term can be computed from a self-consistent QM/MM calculation, followed by a non-self-consistent calculation without external point charges. Alternatively, the term can be computed approximately by subtracting the MM-QM electrostatic interaction energy computed classically from the QM/MM energy.[25] Here, we make use of the former procedure.

QM calculations

The reaction pathway was obtained from QM/MM calculations, in which the QM region was relaxed orthogonal to the reaction coordinate, while the MM region was fixed at the crystal structure.[27] The distance between the acceptor oxygen atom of the catecholate and the carbon atom of the transferred methyl group ($d_{\mathrm{C-}O}$) was used as reaction coordinate, and optimizations of the QM region was performed at $d_{\mathrm{C-}O}$ = 1.47, 1.80, 1.95, 2.00, 2.05, 2.10, 2.13, 2.30, 2.55, and 2.84 Å, altogether 10 points. The calculations were performed with the QM/MM software ComQum,[33,34] which in turn is based on the molecular mechanics software Amber[35] and the quantum mechanics software Turbomole.[36] Density functional theory with the PBE exchange-correlation functional[37] and the 6-31G* basis set, extended with diffuse basis functions on nitrogen, oxygen, and sulfur, was used to describe the QM region, whereas the MM region was described by the Amber94 force field.[38]

For each point on the reaction pathway, restrained electrostatic potential (RESP) point

charges were calculated with Amber.[39] These calculations were in turn based on calculations of the electrostatic potential on a grid using the Merz-Kollman scheme[40] using the Gaussian98 software package.[41]

The single-point quantum mechanical energies in curled brackets in Eq. 6 were calculated using the hybrid density functional Becke-3-Lee-Yang-Parr (B3LYP) functional as implemented in Turbomole[42] and the 6-311++G(2d,2p) basis set. Strictly speaking, we should use the same level of methodology as employed to compute the point charges, but we have previously shown that extrapolation to another set of exchange-correlation functional and basis set is possible to a high degree of accuracy.[25,26]

Simulations

For each point along the reaction pathway, a simulation was performed for 600 ps using either explicit or implicit solvent. The last 400 ps were used to collect data for the subsequent free energy calculations. In all simulations, a time step of 2 ps was employed and the SHAKE algorithm[43] was used to constrain the lengths of bonds containing hydrogen bonds. Moreover, all atoms in the QM region were fixed and described by point charges obtained from the QM/MM calculations.

The explicit solvent simulations have been reported previously[25,26] and we refer to those articles for more details than given here. The simulations were performed with periodic boundary conditions using an octahedral box and a constant volume. The temperature was kept constant by reassigning the velocities if the temperature deviated by more than 5 K from 298 K. About 7800 water molecules described by the TIP3P model were used in the simulation of the enzyme-catalyzed reaction, whereas about 8800 molecules were

used for the reaction in solution. Electrostatic interactions were calculated with the particle mesh Ewald method.[44]

In the implicit solvent simulations, the generalized Born molecular volume (GBMV) method was used to describe the solvent. The atomic radii are parameters in the equation for the Born radii, but they also define the volume and surface of the solute. However, despite the importance of the atomic radii, they cannot be unambiguously defined and different sets of radii have been suggested. The calculations were made for 4 different atomic radii, namely the radii proposed by Roux and coworkers[45,46] (Set I), the van der Waals radii defined by the Amber94 force field[39] (Set II). Set I scaled by a factor 0.98[47] (Set III), and Set II scaled by 0.87 (Set IV). For all sets of radii, we used an angular integration grid with $N_\phi = 8,$ but for Set II we also tested a coarser integration grid with $N_\phi = 5.$ A surface tension of 0.005 kcal/(mol Å2) was used for the non-polar term. The implicit solvent simulations were started from structures that were minimized by 500 steps using the adopted-basis Newton-Raphson methodology. The temperature was kept constant at 298 K by employing Langevin dynamics using a friction coefficient of 5 ps-1 applied to all heavy atoms.

Because the routine that fix atoms in CHARMM, modifies the non-bonded list, it is not straightforward to fix atoms in combination with GBMV. Under the kind guidance from Michael S. Lee, we modified the CHARMM code, so that we can fix atoms without eliminating any electrostatic terms. Unfortunately, this works only if we use no cutoff, and, hence, the implicit solvent simulations were performed with an infinite cutoff. Consequently, the simulations were very slow and it does not make sense to compare timings of the explicit and implicit simulations.

For the reaction in solution, all atoms of the QM region were fixed so the implicit solvent simulations reduce to single-point calculations. In order to avoid inaccuracies owing to the integration grid, we used a very fine grid ($N_{\phi}=12$) and report the average of 9 different calculations, which differ in the orientation of the solute molecule with respect to the grid.

Analysis

The simulations were analyzed with respect to root-mean-square deviations (RMSD) from the crystal structure. Moreover, the number of hydrogen bonds and salt bridges were calculated for each simulation. We define a hydrogen bond as a contact between a hydrogen atom and a different atom for which the inter-atomic distance is less than 2.4 Å at least 66% of the time during the last 400 ps of the simulation. A salt bridge is defined as a hydrogen bond where the involved hydrogen atom is one of the hydrogen atoms in the charged groups of Arg, Lys, or in a doubly protonated His, and where the heavy atom is one of the oxygen atoms in the side chains of Asp or Glu.

Free energy calculations

To perform the free energy calculations, 400 configurations evenly spaced in time and taken from the last 400 ps of simulation were stored. For each of the configurations, the non-perturbed and perturbed energies were calculated. These energies were used to compute the free energy change between the non-perturbed and perturbed state by means of free energy perturbation. Recent studies have shown that Bennet's acceptance ratio method[48] is more reliable than free energy perturbation,[49,50] but test calculations showed

that the effect is insignificant for the free energy changes studied here.

Results

The methyl transfer reaction

Before we consider the different solvent models, we will discuss results based on the explicit solvent models and compare the methyl transfer in solution with the corresponding enzyme-catalyzed reaction. In the left panel of Figure 2, the free energy is decomposed into contributions from the B3LYP/6-311++G(2d,2p) calculations of the QM region and the classical interaction free energy between the QM and MM region, cf. Eq. 6. The QM energies show that the product complex (i.e. neutral methylcathecol and S(CH3)2 at d_{C-O} = 1.47) is strongly favored in a vacuum. This is because the reaction neutralizes the charges of the reactants (catecholate and S(CH3)3+). However, in the enzyme, this stabilization is less pronounced, because the Mg2+ ion and its ligands destabilize the product complex relative to the reactant complex (note that the configurations of atoms common for the QM systems used to model the reaction in solution and catalyzed by the enzyme are identical, cf. Figure 1). This is partly because some of the charge of the catecholate is transferred to the Mg2+ ion (and its ligands; ~0.06 e), and partly because a sizable dipole moment is preserved after the reaction in the presence of the Mg2+ ion. The dipole moment of the QM region changes from 22 to 3 Debye for the reaction in solution, whereas it changes from 24 to 16 Debye when the Mg2+ complex is included in the QM region.

For each of the two reactions, results are given for the MM-QM interaction free energies

based on forward as well as reverse free energy perturbation calculations. The hysteresis amounts to less that 2 kJ/mol in both cases. The interaction free energies show that the environment modeled by the MM atoms destabilizes the product complex relatively to the reactant complex. This can again be understood in terms of the charge neutralization reaction, because a medium with a higher dielectric constant favors the polarized reactant complex more than the less polarized product complex. For the same reason, it is expected that the MM environment inhibits the reaction in solution more than the enzyme-catalyzed reaction.

In the right panel, the total energy (sum of QM energy and MM-QM interaction free energy) are plotted. The barrier for the enzyme-catalyzed reaction is 68 kJ/mol and that of the reaction in solution along the same reaction pathway is slightly lower, namely 55 kJ/mol. We note that the minimum free energy path for the reaction in solution probably differs from the barrier computed here, partly because the reaction pathway will be different and partly because the end points utilized here are not minima. Moreover, for a full comparison between the reaction in solution and the enzyme-catalyzed reaction, the binding affinity of the reactant molecules should also be considered. It is also possible to have a higher barrier for the enzyme-catalyzed reaction than for the corresponding reaction in solution, because the methyl transfer presumably is preceded by a deprotonation step, which is unfavorable in solution, owing to the high pKa value of the catechol (9.5).[51] In the enzyme, the pKa value is reduced, owing to interactions with the positively charged Mg2+ ion. This is probably one of the most important mechanisms by which COMT catalyzes this reaction. For a discussion of catalytic factors, see also Roca et al.[52]

Reaction in solution

We start by reporting results for the calculations for the small system defined in Figure 1b. In the left panel of Figure 3, the free energy for the reaction in solution are plotted for the explicit TIP3P solvent simulations and GB calculations based on four different sets of Born radii, as explained in the Methods section.

It can be seen that the Amber van der Waals radii (set II) give energies that are more similar to those obtained by an explicit solvent than the Born radii suggested by Roux and coworkers (set I). The scaling of 0.98 has only a minor influence on the results (set III). Not surprisingly, the radii of the atoms that are directly involved in the reaction affect the results significantly, and by tuning the radii of those atoms, an almost perfect match between the explicit solvent and the implicit solvent results can be obtained, as observed with Set IV. The same agreement cannot be obtained by tuning the radii of the central atoms in Set I, because the radii of the methyl hydrogen atoms in that set are zero. The deviation between the implicit and explicit solvent is most noticeable at and around the transition state, whereas the curves match quite well at the end points. This can also be seen from Table 1 where the activation and reaction free energies are listed for the different solvent models. This is not surprising, because the atoms change their physical properties along the reaction, and in a force field, it therefore makes sense to let the parameters for the central atoms change along the reaction pathway. Such a procedure is shown in the right panel of Figure 3, in which a curve with the radii of central atoms scaled by 0.71 is compared to two other curves which are based on calculations with the radii of the central atoms scaled by 0.90. Individual offsets are added to the curves, so

that they match the one based on explicit solvent simulations at 2.05 Å, and at each of the two end points, respectively. As can be seen, all the curves follows the curve based on explicit solvent simulations almost perfectly around the point where the implicit and explicit solvent curves have been matched.

Enzyme reaction

The protein turns is stable in all simulations performed, with fairly small deviations from the crystal structure. However, the root-mean-square-deviations from the crystal structure are in general larger in the implicit solvent simulations than in the explicit simulations. A typical example is shown in Figure 4 for a simulation of the transition state complex, and a summary is listed in Table 2. The average deviations are normally between 0.6 and 0.8 Å for the Cα atoms in the explicit solvent simulations and between 0.8 and 1.1 Å in the implicit solvent simulations if Set III is used as Born radii, whereas they are between 1.0 and 1.5 if Set II is used. Hence, Set II performs better in comparison with the crystal structure and explicit solvent.

Another descriptor that deviates between the implicit and explicit solvent simulations is the number of hydrogen bonds, cf. Table 3. Based on the definition of hydrogen bonds given in the Method section, 80% of the hydrogen bonds in the explicit solvent simulations are preserved in a corresponding implicit solvent simulation. Nonetheless, the total number of hydrogen bonds is 5-20% higher in the implicit solvent simulations than in the corresponding explicit solvent simulation.

Problems with salt bridges have been reported previously and here we find a similar trend for salt bridges as we found for the hydrogen bonds: There are more salt bridges in the

implicit solvent simulations, but not all salt bridges in the explicit solvent simulations are preserved in the corresponding implicit solvent simulation.

In conclusion, we find that the implicit solvent simulations are robust and stable, but that there also are some structural changes between the implicit and explicit solvent simulations. Related to this study is a comparative study of different GB models by Fan et al.[53] They found deviations between number of hydrogen bonds and radius of gyration between different GB models and between GB models and experimentally determined structures.

Free Energy Calculations

We compare free energy barriers based on explicit and implicit solvent models and combination thereof. Again we compare different sets of atomic radii, namely Set II and Set III as defined in the previous section. For both sets, we have made simulations for each point along the reaction pathway followed by free energy perturbations. We have also used a procedure where we reprocess the explicit solvent simulations with an implicit solvent model using either Set II or III as described in the Method section. We refer to these calculations as TIP3P/GBMV(II) or TIP3P/GBMV(III). In general, we use a fine integration grid ($N_\phi=8$; see Method section) but for Set II we also investigate a coarser integration grid $(N_\phi=5)$ and the possibility of performing free energy perturbation calculations by reprocessing simulations based on the coarser integration grid with the finer integration grid, referred to as GBMV(II,5/8).

Computed reaction and activation free energies are listed in Table 4 and some of the free energy curves are plotted in Figure 5. In Figure 6, the percentage deviation of the MM-QM interaction free energy from that of the explicit solvent simulations are plotted for

the forward and reverse activation energies as well as for the reaction energy. It can be seen that the results based on the coarser integration grid perform poorly compared to results based on the finer integration grid. Overall, GBMV(III,8) gives results closest to those with explicit solvent, but the deviation between the results based on Sets II and III are small, less than 1 kJ/mole for the three free energy changes listed in Table 4. The same trend is observed for the TIP3P/GBMV results, cf. Table 4. Hence, as opposed to the reaction in solution, the free energies for the enzyme-catalyzed reaction are fairly robust towards the choice of atomic radii. For the same reason, we assume that using the unscaled Set I rather than Set III would not change the results significantly.

On the other hand, we find that the results are not converged with respect to the integration grid if $N_\phi=5$ is used. In Figure 5, the free energy curves based on the explicit solvent simulations are compared to results based on the implicit solvent simulations using Set III but with different integration grids. It is seen that the coarser grid gives a higher barrier for the forward reaction and a lower barrier for the reverse reaction than obtained with any of the other methods. Reprocessing the simulations with the finer integration grid does not correct the errors and therefore does not seem to be a valid strategy. Moreover, the energies evaluated with the two integration grids differ too much to make a free energy perturbation correction possible.

Despite the agreement between results based on Set II and Set III, hysteresis effects are larger with Set II. Barriers computed stepwise and forwardly or reversely, deviate at most by 3 kJ/mol when using TIP3P/GBMV(III,8) or GBMV(III), but they may deviate as much as 8 kJ/mol when using Set II, i.e., the Amber van der Waals radii. The effect of 8 kJ/mol is comparable to the hysteresis effect of using Set III with a course integration

grid. On top of that, Set II is more expensive to employ than Set III, because Born radii need to be computed for all atoms with a non-zero van der Waals radius, whereas the atomic radius is set to zero for all hydrogen atoms in Set III. Moreover, as previous seen, the RMSD from the crystal structure is larger for Set II than III.

To conclude this section, we find that the implicit solvent methods perform reasonably well when compared to results based on explicit solvent simulations, particularly if a proper integration grid is used.

Single-stage free energy calculations for large perturbations

Since the implicit solvent relaxes instantaneously to a modification of the solute, it is possible that the implicit solvent performs better than the explicit solvent if free energy calculations are to be performed over large perturbations in a single step. This is investigated in Figure 7, where deviations of free energy curves based on single simulations of the product or reactant complex are plotted. The deviations are with respect to corresponding ones based on multi-stage free energy calculations, cf. Figure 5. Not surprisingly, it is seen that the deviations increase with larger steps. As expected, the deviations observed when using explicit solvent diminish if the energies are reprocessed with implicit solvent (TIP3P/GBMV(III,8)). The results based on implicit solvent simulations (GBMV(III,8)) are a bit surprising because free energy calculations based on a simulation of the product complex is capable to reproduce the entire free energy barrier within 2 kJ/mol (blue dashed line with diamonds) and thus performs best among the three methods under consideration. In contrast, an implicit solvent simulation of the reactant complex is the poorest way to generate the entire free energy curve (red solid line with

diamonds).

For the current reaction, there is a limit of how much the implicit solvent model can account for the polarization of the MM environment, since most of the change in polarization occurs in the protein matrix rather than in the solvent. This can be seen from Figure 8, where the electrostatic potential on the methyl donor and acceptor atoms caused by the solvent and the MM protein environment are plotted and compared with the reaction in solution. As can be seen, there is a gap between the distributions of the product and reactant complex for the part originating from the protein, whereas there is a significant overlap between the distributions of the solvent. Hence, the protein is polarized differently in the product and reactant complexes, whereas that is less pronounced for the solvent. In contrast, there is a significant gap between the distributions for the reaction in solution.

A gap between the distributions of the (explicit) solvent degrees of freedom does not pose a problem for an implicit solvent model, because it will be solved by the instantaneous relaxation. However, this is not the case for the protein degrees of freedom, and the insignificant overlap between the distributions originating from the protein therefore pose a problem in both the implicit and the explicit solvent simulations, although the effect of a wrongly polarized protein will be somewhat dampened by an implicit solvent model. Therefore, we cannot expect a good agreement between the multi-stage and the single-stage free energy calculations over the entire reaction pathway for the enzymatic reaction.

Conclusions

In this paper, we have considered the methyl transfer reaction between S-

adenosylmethionine and catechol, catalyzed by the COMT enzyme. We have seen that in vacuum, the products are strongly favored, because they are neutral. However, both the Mg2+ ion and solvation effects instead favors the charged reactants. As a result, the activation energy of the reaction in the enzyme and in solution is quite similar, 68 and 55 kJ/mole, respectively (but the reaction is more exothermic in solution). However, this disregards the role of deprotonating the catechol molecule, and possible differences in reaction path between the enzyme and solution.

However, the main goal of this article has been to compare the performance of the implicit GBMV model with explicit solvent models for the calculation of free energies. We have seen that for the COMT reaction in solution, the resulting free energies are quite sensitive to the parameters of the method (e.g. the atomic radii). In this case, the Amber van der Waals atomic radii (Set II) performed better than those suggested by Roux et al. (Set I). Moreover, results very close to those with explicit solvent could be obtained by simply scaling the parameters, especially for the atoms involved in the chemical reaction. In the enzyme, structural parameters (e.g. the RMSD from the crystal structure, or the formation of hydrogen bonds and salt bridges) are still quite sensitive to the atomic radii. Moreover, there are pronounced differences between the calculations with implicit and explicit solvent. In this case, the (scaled) Roux atomic radii (Set III) gave results closer to those of the explicit solvent simulations.

However, these structural differences have little effect on the activation and reaction energies: The GBMV calculations reproduce the activation and reaction energies in the explicit solvent simulations within 5 kJ/mole and there is little difference between the two sets of atomic radii (set II and III).

We have also examined if it possible to calculate the activation and reaction energies from a single MD simulation, employing the instantaneous relaxation of the solvent degrees of freedom in the implicit methods. This does not seem to be possible, because the main difference in the polarization of the environment takes place in the protein, rather than in the solvent, and there is little overlap in the protein polarization between the reactant and product complexes (cf. Figure 8).

Thus, the studies of COMT suggest that the GBMV model performs well in free energy calculations if detailed solute-solvent interactions are not important. Therefore, it works better for enzymatic reactions than for reactions in solution. On the other hand, there is less gain from the GB method in studying enzyme catalysis, because the major polarization of the environment takes place in the protein and not in the solvent. Free energy calculations based on explicit solvent and for large fluctuations can be improved by reprocessing trajectories with implicit solvent. Such a procedure has previously been employed by Rod & Brooks.[14] The more solvent exposed the reaction is, the more is gained by such a procedure.

Finally, our results suggest that the atomic radii of Roux et al. work best for enzymatic reactions, whereas the Amber radii works better in solution. However, it is even more important to use a proper integration grid.

Acknowledgments

This investigation has been supported by funding from the Swedish Research Council, and by computer resources by the Swedish National Infrastructure for Computing, Lunarc at Lund University, and the High Performance Computing Center North (HPC2N) at the

University of Umeå. We acknowledge fruitful discussions with Michael Feig, Michigan State University, and the kind help by Michael S. Lee to run simulations with GBMV and fixed atoms.

Tables

Method	ΔA_{reac}	$\Delta A^{\neq}_{\mathrm{fwd}}$	$\Delta A^{\neq}_{\mathrm{rev}}$	d_{TS} ***(Å)***
TIP3P	-88(0.)	55(0.)	143(0.)	2.10
GBMV(I)	-112(18)	46(19)	158(16)	2.13
GBMV(II)	-98(8)	53(11)	152(0)	2.13
GBMV(III)	-110(17)	47(18)	157(13)	2.13
GBMV(IV)	-87(1)	58(5)	145(15)	2.13
QM(vac)	-221(100)	0(100)	221(100)	2.84

Table 1.Reaction and activation free energies for forward and reverse methylation reaction in aqueous solution. Percentage deviations of the MM-QM interaction free energies from those of corresponding explicit solvent (TIP3P) simulations are given in parentheses.

ρ_i ***(Å)***	***TIP3P***	***GBMV(II,8)***	***GBMV(III,5)***	***GBMV(III,8)***
1.47	*0.7*	*1.4*	1.0	0.9
1.80	0.7	*1.2*	1.0	0.9
1.95	0.7	*1.3*	1.1	1.0
2.00	0.7	*1.5*	0.8	0.9
2.05	0.6	*1.3*	0.9	0.9
2.10	0.7	*1.3*	0.9	1.1
2.13	0.7	*1.2*	1.0	0.9
2.30	0.7	*1.5*	1.1	1.0
2.55	0.7	*1.0*	1.1	1.0
2.84	0.7	*1.6*	1.1	1.0

Table 2. Average root-mean-square-deviations of Cα atoms from crystal structure.

ρ_i (Å)	*TIP3P*	*GBMV(III,5)*	*GBMV(III,8)*	*GBMV(II,8)*
1.47	106	112/95	109/97	109/90
1.80	106	113/94	120/99	107/94
1.95	107	112/92	118/101	107/89
2.00	109	112/96	110/93	106/92
2.05	101	112/88	113/94	112/90
2.10	107	115/97	117/99	105/93
2.13	108	115/99	119/102	109/97
2.30	108	111/93	123/100	104/91
2.55	108	113/92	112/97	112/94
2.84	108	112/95	114/99	106/97
Preserved	87	84/71	93/78	86/74

Table 3. Number of hydrogen bonds between backbone atoms (O and N) in simulations of the different configurations of the QM region. The numbers after the slashes indicate the number of hydrogen bonds common with the corresponding explicit solvent (TIP3P) simulation. The number of hydrogen bonds that are preserved in all simulations in a column, are listed in the last row.

Method	ΔA_{reac}	$\Delta A^{\neq}_{\text{fwd}}$	$\Delta A^{\neq}_{\text{rev}}$	d_{TS} ***(Å)***
TIP3P	-25(0.)	68(0.)	92(0.)	2.05
TIP3P/GBMV(II,8)	-17(10)	71(7)	87(15)	2.05
TIP3P/GBMV(III,8)	-19(7)	70(4)	89(11)	2.05
TIP3P/GBMV(IV,8)	-17(10)	71(7)	88(15)	2.05
GBMV(II,8)	-30(7)	65(6)	95(9)	2.05
GBMV(III,8)	-29(6)	65(7)	94(5)	2.05
GBMV(III,5)	-12(16)	74(15)	86(19)	2.05
GBMV(III,5/8)	-35(14)	61(25)	96(2)	2.10
QM/MM	-7(18)	68(0)	75(45)	2.05
QM(vac)	-98(100.)	27(100.)	125(100.)	2.13

Table 4.Reaction and activation free energies for the methyl transfer reaction catalyzed by COMT. The values in parentheses are percentage deviation of the MM-QM interaction free energy from that of the corresponding explicit solvent (TIP3P) simulations.

Figure Captions

Figure 1.Atoms in the quantum system in the calculations with (a) or without (b) the enzyme (SAM=S-adenosylmethionine).

Figure 2. The left panel shows the contributions to the total QM/MM free energy from the QM region and the MM-QM interaction free energy for the enzyme-catalyzed reaction (E) and the reaction in solution (S). Right panel shows the total QM/MM free energy for the two reactions.

Figure 3. Free energy barriers computed for the methylation reaction in aqueous solution. In the left panel, barriers computed with explicit (TIP3P) and implicit. (GBMV) solvent using different Born radii (I-IV) are compared. In the right panel, barriers computed with the implicit solvent model and usingthe set II radii scaled by either 0.7 or 0.9 are shown. The barriers are matched to either the reactant complex (RC), transition state complex (TSC), or the product complex (PC).

Figure 4.Root-mean-square-deviations (RMSD) for C_α atoms for a simulation of the transition state complex. Only the last 400 ps are shown..

Figure 5.Computed free energy barriers for the enzyme-catalyzed methylation reaction. Different GB implicit solvent models are compared with explicit solvent (TIP3P).

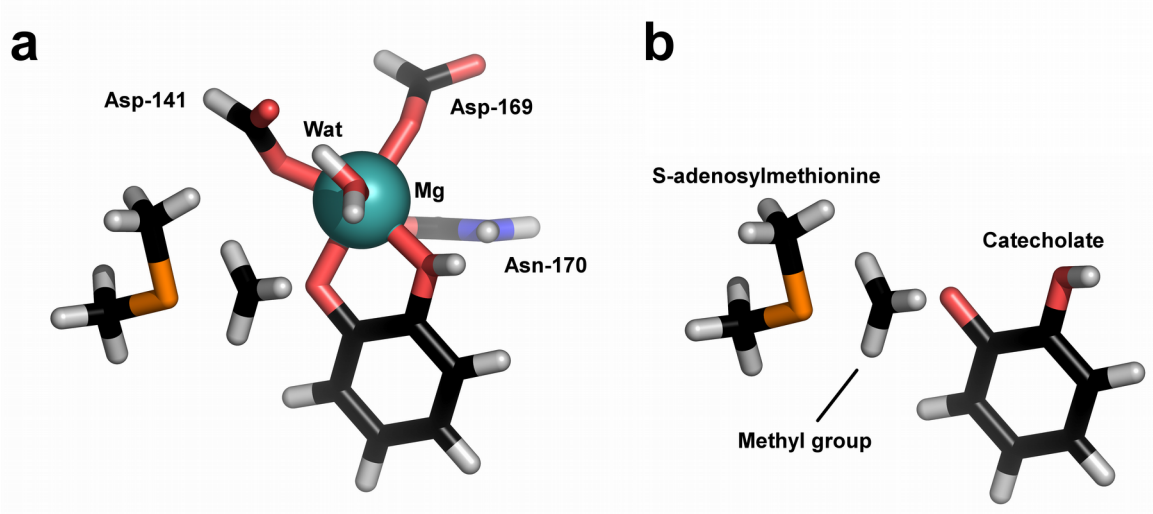
Figure 6.Relative deviation of MM-QM interaction free energy changes from those

computed with an explicit solvent model (TIP3P). Red (left most) bars show deviations for the reaction energy, whereas green (center) and blue (right most) bars show deviations for the activation energy of the forward and reverse reactions, respectively.

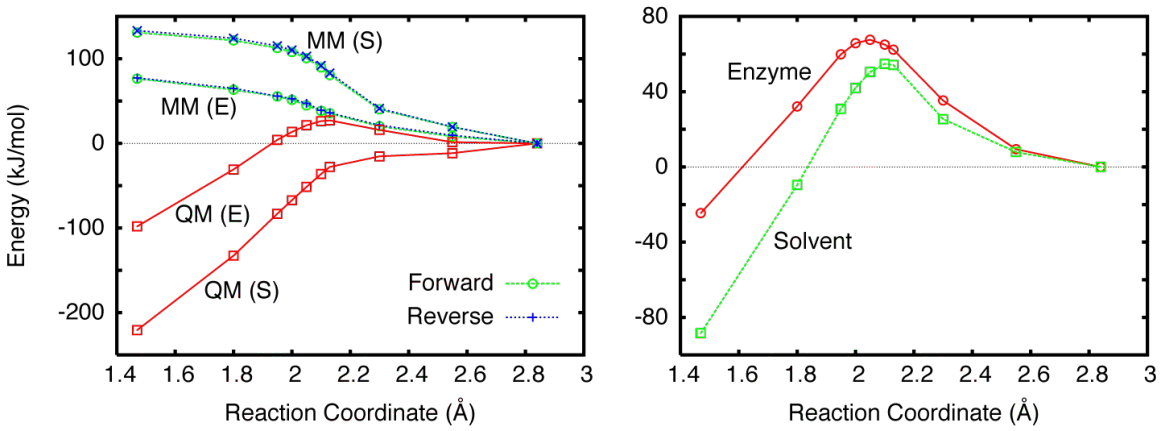
Figure 7. Deviations of free energy barriers, which are computed by single step free energy perturbation, from the corresponding ones, that are computed by a multi-stage free energy perturbation calculations, cf. Figure 5. Red solid lines are for calculations based on simulations of the reactant complex, whereas the blue dashed lines are based on simulations of the product complex. Three different solvent models are employed, namely explicit TIP3P solvent (TP, boxes), implicit GBMV(III,8) solvent (GB, circles), and explicit solvent reprocessed with GBMV(III,8) (TP+GB, diamonds).

Figure 8.Contributions from protein and solvent to the electrostatic potential on the acceptor oxygen atom (O2) versus the potential on the donor sulfur (SD) atom for the reactant complex (RC) and the protein complex (PC). The two broad distributions are for the reaction in the solution, whereas the two pairs of narrower distributions are for the enzymatic reaction separated in to contributions from the protein (lower left) and from solvent (lower right).

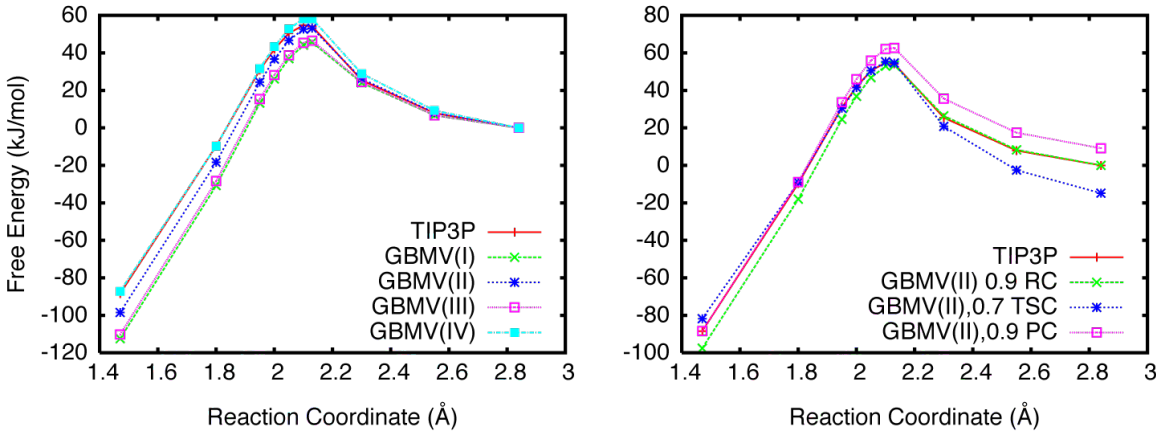
Thomas H. Rod et. al. Figure 1.



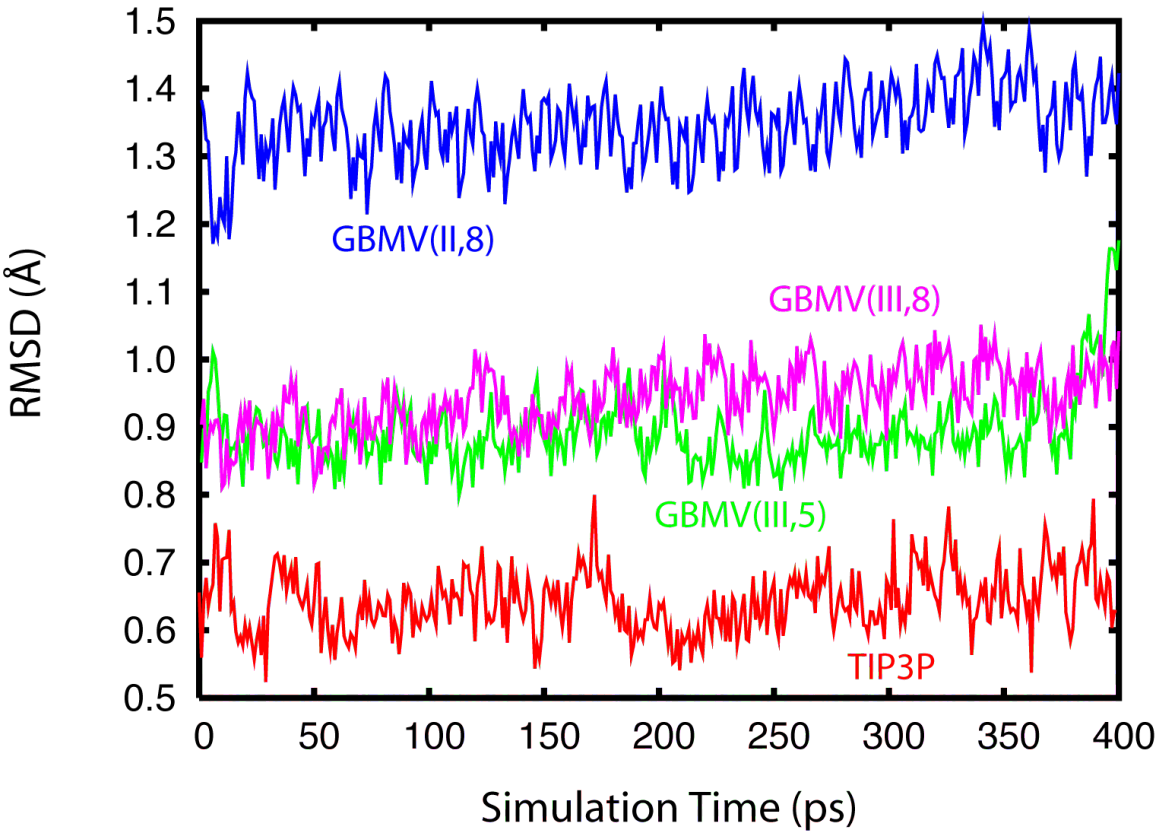
Thomas H. Rod et. al. Figure 2.



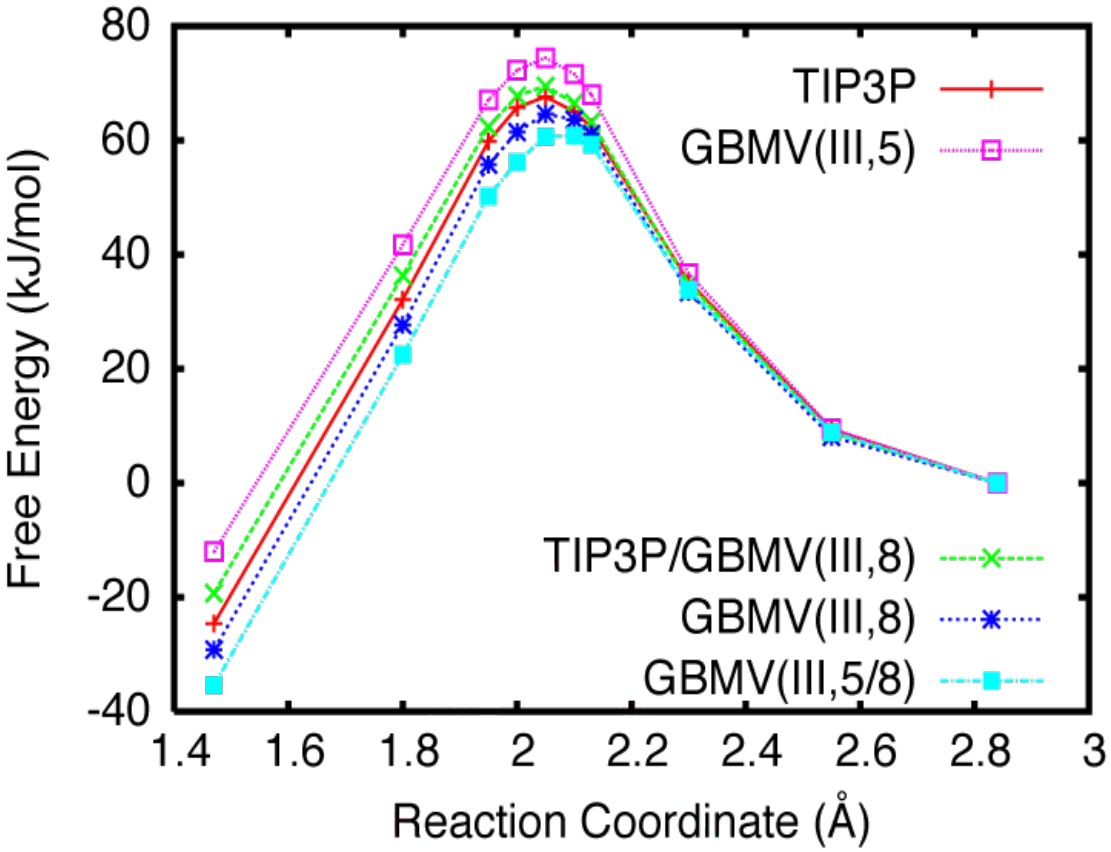
Thomas H. Rod et. al. Figure 3.



Thomas H. Rod et. al. Figure 4.

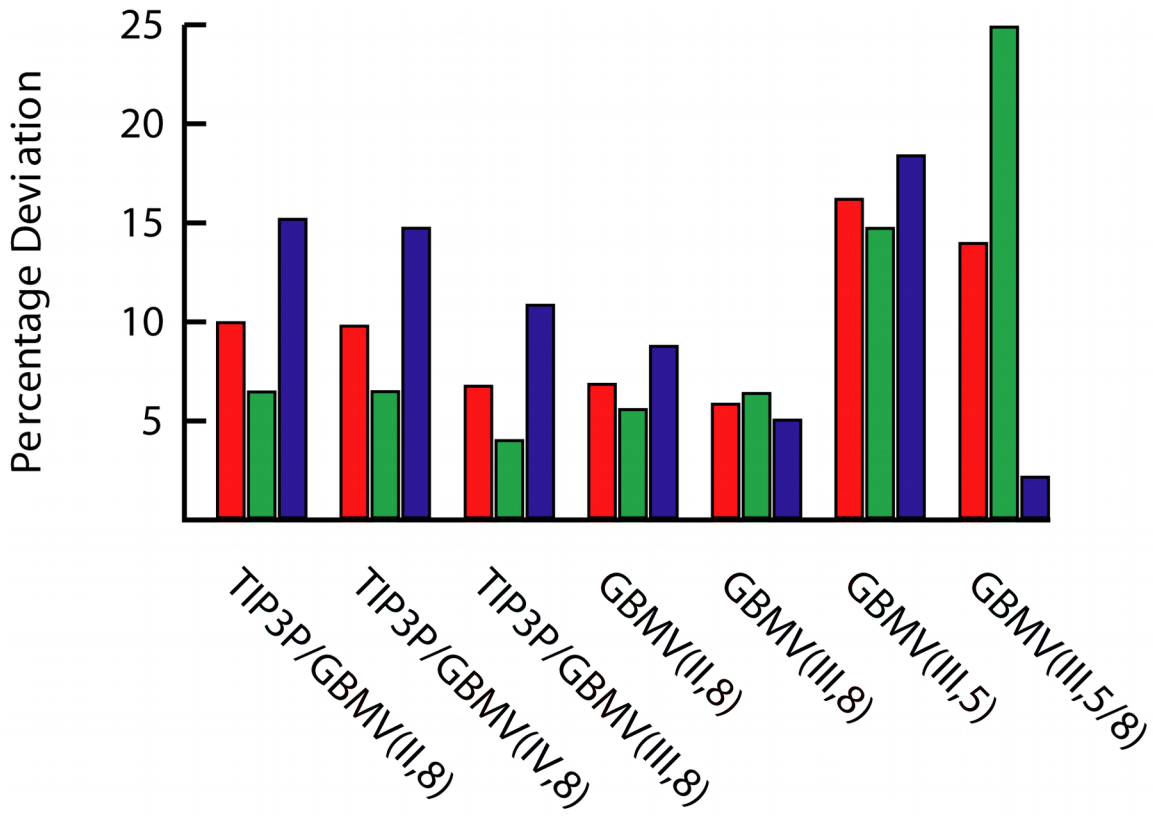


Thomas H. Rod et. al. Figure 5.

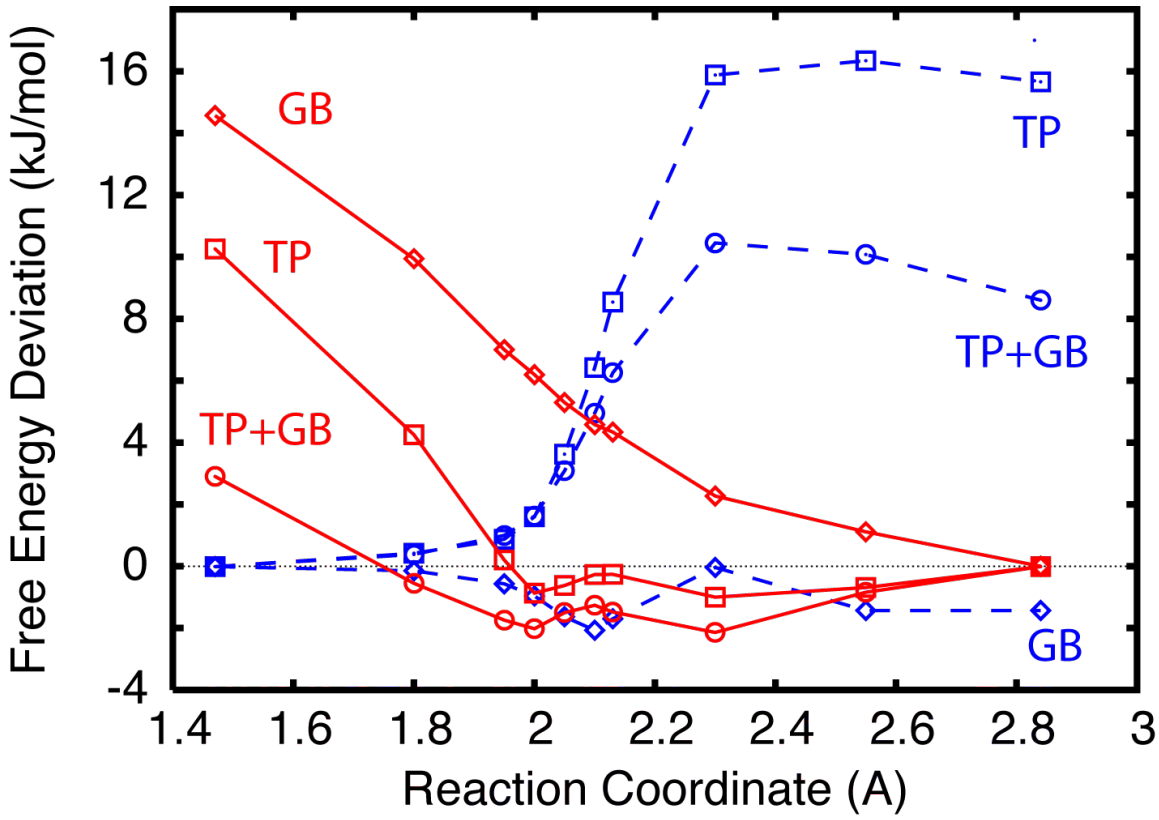


Thomas H. Rod et. al. Figure 6

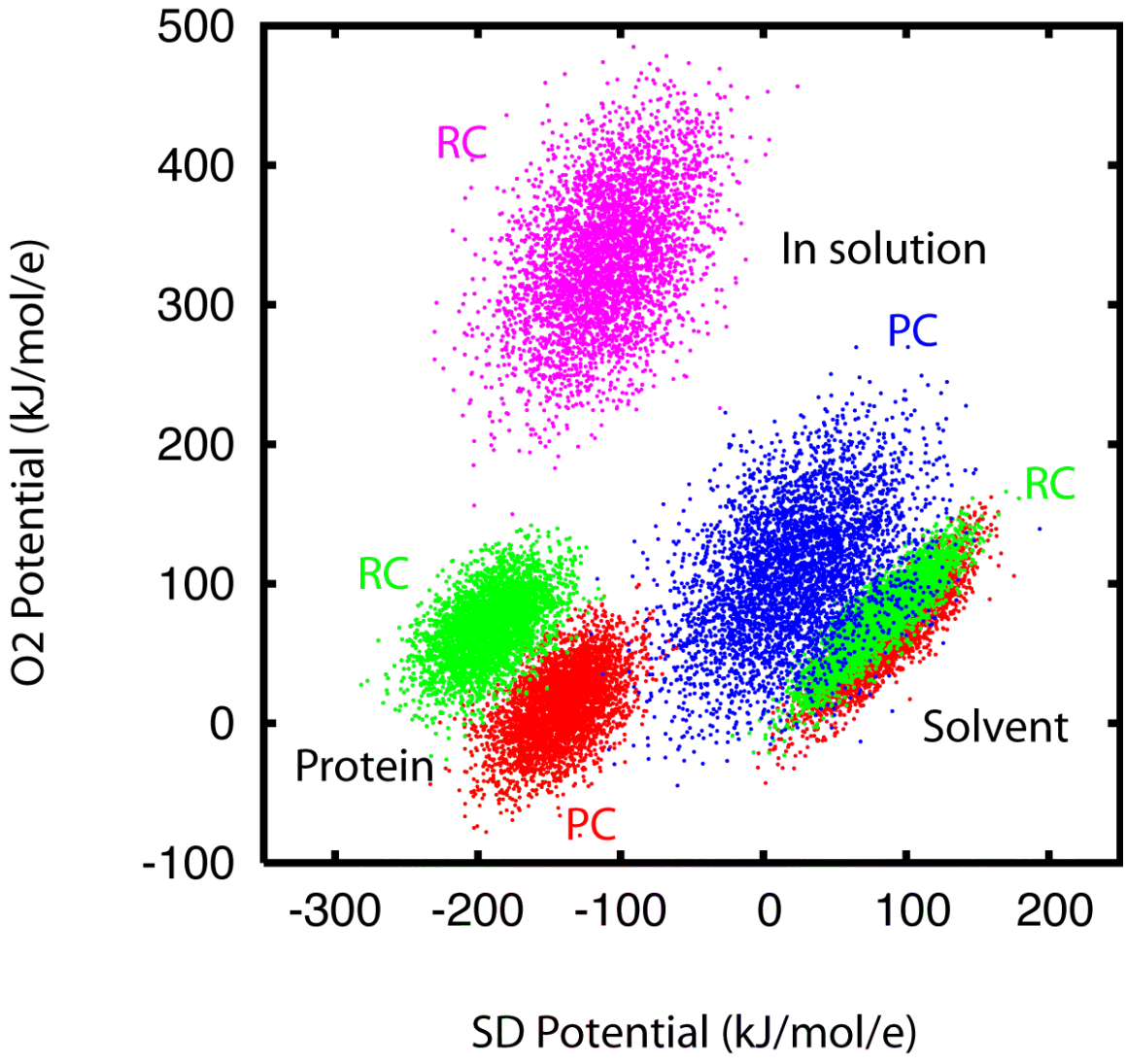
.



Thomas H. Rod et. al. Figure 7



Thomas H. Rod et. al. Figure 8



[1] M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A. Case, and C. L. Brooks III, J. Comp. Chem. **25**, 265 (2003).

[2] W. C. Still, A. Tempczyk, R. D. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).

[3] A. Onufriev, D. A. Case, and D. Bashford, J. Comp. Chem., **23**, 1297 (2002).

[4] J. Zhu, E. Alexov, and B. Honig, J. Phys. Chem. B **109**, 3008 (2005).

[5] H. Fan, A. E. Mark, J. Zhu, and B. Honig, Proc. Natl. Acad. Sci. USA **102**, 6760 (2005).

[6] T. Simonson, J. Carlsson, and D. A. Case, J. Am. Chem. Soc. **126**, 4167 (2004).

[7] R. Zhou, Proteins **53**, 148 (2003).

[8] C. M. Stultz, J. Phys. Chem. B **108**, 16525 (2004).

[9] R. Zhou and B. J. Berne, Proc. Natl. Acad. Sci. USA **99**, 12777 (2002).

[10] H. Nymeyer and A. E. Garcia, Proc. Natl. Acad. Sci. USA **100**, 13934 (2003).

[11] B. D. Bursulaya and C. L. Brooks, III, J. Phys. Chem. B **104**, 12378 (2000).

[12] M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).

[13] A. Onufriev, D. Bashford, and D. Case, J. Phys. Chem. B **104**, 3712 (2000).

[14] T. H. Rod and C. L. Brooks, III, J. Am. Chem. Soc. **125**, 8718 (2003).

[15] H. Liu, Y. Zhang, and W. Yang, J. Am. Chem. Soc. **122**, 65560 (2000).

[16] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T.E. Cheatham, Acc. Chem. Res. **33**, 889 (2000).

[17] J. Khandogin and C. L. Brooks, III, Biophys J. **89**, 141 (2005).

[18] M. S. Lee, F. R. Salsbury, Jr., and C.L. Brooks, III, Proteins, **56**, 738 (2004).

[19] J. Mongan, D.A. Case, and J.A. McCammon, J. Comp. Chem. **25,** 2038 (2004).

[20] J. Mongan and D.A. Case, Curr. Opin. Struct. Biol. **15,** 157 (2005).

[21] Y. Zhang, H. Liu, and W. Yang , J. Chem. Phys. **112**, 3483 (2000).

[22] G. A. Cisneros, H. Liu, Y. Zhang, and W. Yang., J. Am. Chem. Soc. **125**, 10384 (2003).

[23] T. Ishida and S. Kato, J. Am. Chem. Soc. **125**, 12035 (2003).

[24] T. Ishida and S. Kato, J. Am. Chem. Soc. **126**, 7111 (2003).

[25] T. H. Rod and U. Ryde, J. Chem. Theory Comput., **1**, 1240 (2005).

[26] T. H. Rod and U. Ryde Phys. Rev. Lett., **94**, 138302 (2005).

[27] T. Rasmussen, K. Nilsson, and U. Ryde, in preparation.

[28] M. S. Lee, F. R. Salsbury, Jr., and C. L. Brooks, III, J. Chem. Phys. **116**, 10606 (2002).

[29] M. S. Lee, M. Feig, F. R. Salsbury, C. L. Brooks III, J. Comput. Chem. **24**, 1348 (2003).

[30] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[31] J. Vidgren, L. A. Svensson, and A. Liljas, Nature **368**, 354 (1994).

[32] N. Reuter, A. Dejaegere, B. Maigret, M. Karplus, J. Phys. Chem. A **104**, 1720 (2000).

[33] U. Ryde, J. Comput.-Aided Mol. Design **10**, 153 (1996).

[34] U. Ryde, M. H. M. Olsson, Intern. J. Quant. Chem., **81**, 335 (2001).

[35] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, B. Wang, D. A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui. P. Beroza, C. Schafmeister, J. W. Caldwell, W. S. Ross, and P. A Kolman (2004) AMBER 8, University of California, San Francisco.

[36] O. Treutler and R. Ahlrichs, J. Chem. Phys. *102*, 346 (1995).

[37] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996)

[38] W. D. Cornell, P. I. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, **J. Am. Chem. Soc.**, **117**, 5179 (1995).

[39] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, J. Phys. Chem. **97**, 10269 (1993).

[40] B. H. Besler, K. M. Merz, P. A. Kollman, J. Comput. Chem. **11**, 431 (1990).

[41] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Knudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, and J. A. Pople 1998. Gaussian 98, Revision A.9, Gaussian, Inc. Pittsburgh PA.

[42] R. H. Hertwig and W. Koch, Chem. Phys. Lett. **268**, 345 (1997).

[43] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, J. Comput. Phys. **23**, 327-341 (1977).

[44] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).

[45] M. Nina, W. Im and B. Roux, Biophys. Chem. **78,** 89 (1999).

[46] N. K. Banavali and B. Roux, J. Phys. Chem. B **106**, 11026 (2002).

[47] This scaling is historical. Rather than redo all simulations, we have chosen to explore if it has any significance to use the scaling.

[48] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).

[49] N. Lu, J. Singh, and D.A. Kofke, J. Chem. Phys. **118**, 2977 (2003).

[50] Michael R. Shirts & Vijay S. Pande, J. Chem. Phys. **122**, 144107 (2005).

[51] M. Paetzel and R. E. Dalbey. Trends Biochem. Sci. 22, 28 (1997).

[52] M. Roca, S. Marti, J. Andrés, V. Moliner, I. Tuñon, J. Bertrán, and I. H. Williams, J. Am. Chem. Soc. **125**, 7726 (2003).

[53] H. Fan, A. E. Mark, J. Zhu, and B. Honig, Proc. Natl. Acad. Sci. **102**, 6760 (2005).