Quantum mechanical free energy barrier for an enzymatic reaction

Rod, Thomas; Ryde, Ulf

Published in:
Physical Review Letters

DOI:
10.1103/PhysRevLett.94.138302

2005

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

Quantum mechanical free energy barrier for an enzymatic reaction

Thomas H. Rod* and Ulf Ryde
Department of Theoretical Chemistry, Lund University, Chemical Center, P.O. Box 124, S-22100 Lund, Sweden
(Dated: January 4, 2005)

We discuss problems related to in silico studies of enzymes and show that accurate and converged free energy changes for complex chemical reactions can be computed if a method based on a thermodynamic cycle is employed. The method combines the sampling speed of molecular mechanics with the accuracy of a high-level quantum mechanics method. We use the method to compute the free energy barrier for a methyl transfer reaction catalyzed by the enzyme catechol *O*-methyltransferase at the level of density functional theory. The surrounding protein and solvent are found to have a profound effect on the reaction and we show that energies can be extrapolated easily from one basis set and exchange-correlation functional to another. Using this procedure we calculate a barrier of 69 kJ/mol in excellent agreement with the experimental value of 75 kJ/mol.

PACS numbers: 82.20.Wt 82.39.Rt 31.15.-p 82.20.Pm

Enzymes are biocatalysts helping reactions to occur that would not otherwise occur under the mild conditions where life prevails. Enzymes catalyze almost all biochemical reactions and a fundamental understanding of how they function is important for the development of better biocatalysts and drugs against diseases. This becomes increasingly relevant as more genomes are sequenced. Such knowledge is traditionally gathered by experiments but the possibilities to employ computational methods expand rapidly with the continuing gain in computer power and development of efficient quantum and statistical mechanical methods. This development has made in silico design of solid state materials possible from first principle [1], but computer simulations and modeling of soft matter continue to be a demanding task for a number of reasons.

Enzymes catalyze chemical reactions and an accurate description of a chemical reaction demands high-level quantum mechanics (QM) methods capable of describing subtle changes in electron exchange and correlation energy [2]. Advances in density functional theory (DFT) now make it tractable to treat hundreds of atoms at the quantum mechanical level with sufficient accuracy. Unfortunately, the size of nano-particles, macro- and biomolecules exceeds that limit and therefore prevents a full quantum mechanical description of the entire system. The combined quantum mechanics and molecular mechanics (QM/MM) approach circumvents this problem by utilizing that often only a few atoms are directly involved in the chemical reaction, and these are described quantum mechanically, whereas remaining atoms are described by a cheaper molecular mechanics (MM) approach [3, 4]. However, the many degrees of freedom for large and complex systems also imply that sampling of phase space is necessary to obtain accurate thermodynamic properties. Adequate sampling is nonetheless difficult to achieve because of the computational cost of using a high-level QM method on even a modest QM region. Ab initio QM/MM Car-Parrinello like MD simulations are currently limited to a few picoseconds for enzymatic reactions [5, 6], and for that reason, fast but less accurate semi-empirical methods are still widely used because they allow the phase space to be adequately sampled [7–9]. Sometimes the influence from the environment is simply ignored and only the active site is modeled. The protein matrix is, however, essential for a proper understanding of enzymatic reactions and small changes in the protein matrix, in the form of mutations, can have drastic effects [9, 10]. This is illustrated by the enzyme HIV-1 protease, where mutations, even far from the active site, cause drug resistance [5, 10], which is a serious problem in fighting AIDS. In a similar fashion, mutant catechol *O*-methyltransferase enzymes are suspected to be involved in phobic anxiety [11].

Here we compute the free energy barrier for the methyl transfer reaction catalyzed by catechol *O*-methyltransferase (COMT) using a method that combines the sampling speed of molecular mechanics with the accuracy of high-level quantum mechanical methods and that incorporates the effect from the environment in a rigorous way. The method allows for long simulation times (hundreds of picoseconds) and is trivially parallelizable. We study the reaction in solution using a setup of altogether $\sim$27 000 atoms.

The method is based on the thermodynamic cycle depicted in Fig. 1. It shows how a QM/MM (or QM) free energy change between two systems A and B can be calculated as the sum of three terms, namely 1) the free energy difference between A and B described classically, 2) the negative free energy change in going from A described classically to A described by QM/MM, and 3) the free energy change in going from B described classically to B described by QM/MM. The individual free energy changes can be calculated on the basis of simulations by means of free energy perturbation (FEP) by $\Delta F = -kT \ln \langle \exp(-\Delta V/kT) \rangle_{\mathrm{A}}$, where $\Delta V \equiv V_{\mathrm{B}} - V_{\mathrm{A}}$

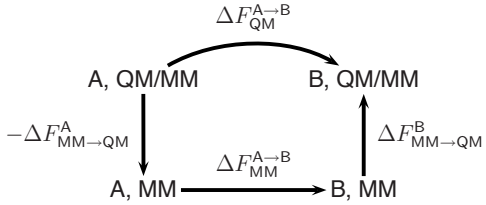

FIG. 1: Thermodynamic cycle used to calculate the QM/MM free energy change, $\Delta F_{\rm QM}^{\rm A\to B}$, between system A and B on the basis of classical sampling of the phase space for system A and B. The free energy change is calculated as $\Delta F_{\rm QM}^{\rm A\to B} = -\Delta F_{\rm MM\to QM}^{\rm A} + \Delta F_{\rm MM}^{\rm A\to B} + \Delta F_{\rm MM\to QM}^{\rm B}$.

and $\langle \ldots \rangle_{\rm A}$ denotes an ensemble average for system A [12, 13].

The two states might be physically different as A and B in Fig. 1 or they might be methodologically different as in Fig. 1 where A (or B) is described either by MM solely or by QM/MM. Notice that by using FEP only one state needs to be sampled in order to calculate the free energy change between two states and therefore a QM/MM free energy change between two states A and B can be calculated by sampling the phase space of A and B *classically* rather than by QM/MM if the thermodynamic cycle in Fig. 1 is employed.

Although QM/MM simulations can be avoided, QM/MM energies should be calculated if FEP is to be applied to the vertical branches in Fig. 1 and, thus, the benefit might seem insignificant. However, in practice when sampling the phase space by either molecular dynamics (MD) or Monte Carlo (MC) simulations, we perform many more energy evaluations than will actually be used in a subsequent calculation of an ensemble average. This is because many simulation steps are needed to move between two points in phase space that are not correlated with one another, i.e. separated by a time longer than the correlation time of the property, in which we are interested [14]. Points separated by less than the correlation time do not provide much additional information. Moreover, many different motions exist in enzymes and some of these can be very slow (in some cases up to seconds) [9] and therefore sampling should generally be performed for as long a time as possible. The thermodynamic cycle approach ensures that quantum mechanical calculations are performed only on points well separated in time, whereas points connecting these points are treated by a cheaper approach. In the simulations for COMT, described below, we perform one QM/MM calculation for every picosecond of simulation, i.e. for every 500 classical MD steps. Compared to the alternative of sampling the QM/MM potential energy surface directly by performing QM/MM MD simulations, this is a significant save, since such simulations demand a QM/MM calculation for every time step, i.e. 500 times as many QM/MM calculations to move the same distance in time as in the thermodynamic cycle approch.

Adapting a thermodynamic cycle to convert a difficult free energy calculation into simpler calculations has been utilized in many studies [13]. Warshel and coworkers have used a cycle similar to Fig. 1 to calculate various free energies [7, 15–17] including an activation energy for an enzyme reaction [7]. However, they obtained poorly converged results because of large fluctuations of the perturbation (~ 40 kJ) and the calculations were always supplemented by results obtained by other methods. In this paper, we base the calculations on optimized QM/MM structures along a reaction coordinate [18] and show that the convergence problem can be avoided by keeping the QM system fixed during the MD simulations.

As a representative enzymatic reaction, we study the methyl transfer reaction catalyzed by the enzyme catechol *O*-methyltransferase (COMT) using DFT. COMT is important in the central nervous system where it inactivates neurotransmitters containing a catechol group, such as dopamine and adrenaline. Neurotransmitters mediate electric signals across synaptic clefts and many diseases are related to malfunction of this part. Patients with Parkinson's disease have low levels of dopamine, which causes uncontrolled muscle movements. Patients are treated with drugs containing levodopa, which in turn is transformed to dopamine in the brain. Levodopa, together with dopamine, are inactivated by COMT and COMT inhibitors are therefore supplied along with levodopa.

In the methyl transfer reaction, a CH_3 group is transfered from a sulfur atom of the cofactor *S*-adenosylmethionine (SAM) to an oxygen atom of catecholate, cf. right panel of Fig. 2 and, hence, a bond is broken and another bond formed. Catecholate binds to the enzyme via a Mg^{2+} ion. The enzyme is relatively small ($\sim$3400 atoms), the transfered group is sufficiently large that tunneling can be ignored, and large-scale motions on longer time scales are presumably absent [19], which makes this enzyme a proper test case. The reaction has in addition been considered by other groups using different computational methodologies [8, 19–21] and with vastly different results. Kohn and Kuhn [19] computed a barrier of 94–103 kJ/mol using the so-called quantum mechanical-free energy method where interactions between the QM region and MM region are treated classically. On the other hand, Roca et al. obtained a barrier of 44 kJ/mol using a semi-empirical (AM1) method and a value of 87 kJ/mol when extrapolated to the level of second order Möller-Plesset (MP2) perturbation theory.

Our setup for the calculations are based on the crystal structure [22] and is illustrated in Fig. 2. Ten points along the reaction pathway are sampled by means of constant temperature MD simulations performed with the program CHARMM [23]. In these simulations atoms in the QM region are represented by point charges, which

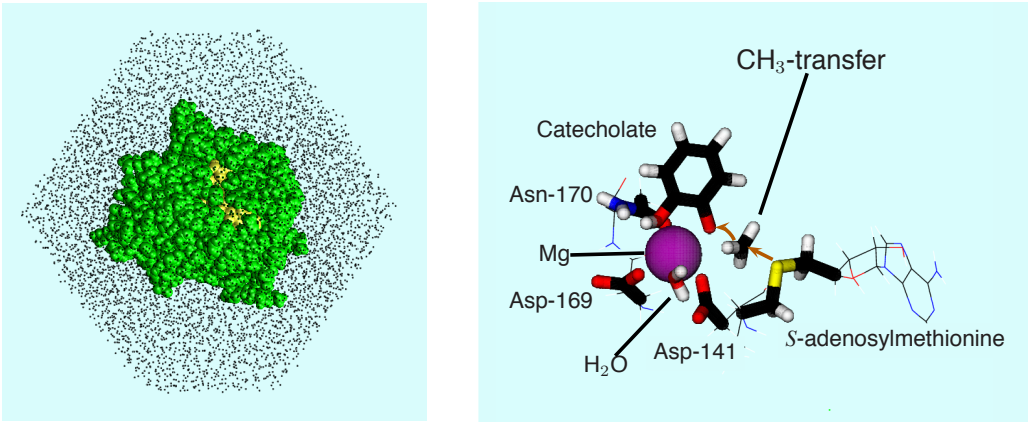

FIG. 2: Setup employed in the simulations. Left panel shows the unit cell employed in the calculations. Solvent molecules are depicted with dots, whereas non-solvent atoms are illustrated by spheres with the active site in a lighter color. Right panel shows a close up of the active site. Atoms in the QM region are depicted with thick sticks and the methyl transfer reaction is indicated with arrows. The QM system consists of the catecholate molecule, the Mg^{2+} ion coordinated by catecholate, an $S(CH_3)_3^+$ molecule to model *S*-adenosylmethionine, $HCOO^-$, $HCOO^-$, and $HCONH_2$ to mimic the Mg^{2+} ligands: Asp-141, Asp-169, and Asn-170, respectively, and a Mg coordinating water molecule.

in turn have been fitted to reproduce the QM electrostatic potential around the active site [24]. Atoms in the QM region are fixed in the simulations, but MM atoms are completely mobile. Solvent is modeled explicitly by the three-point TIP3P model [25] and the protein by the Amber94 force field [26]. Periodic boundary conditions are employed and the particle mesh Ewald method [27] is used to calculate the electrostatic interactions. Each point is simulated for 600 ps, of which the last 400 ps are used to calculate free energy changes. Configurations are stored every 1 ps, resulting in a total of 400 configurations for each simulation. For each of these configurations the energy change upon moving a step along the reaction pathway (horizontal step in Fig. 1, i.e. upon changing point charges and positions of the atoms in the QM region), and by changing from a purely classical description to a QM/MM description (vertical steps in Fig. 1), are calculated and used for the FEP calculations.

The QM system consists of 44 atoms (right panel of Fig. 2). Five of these atoms are so-called link atoms that are bonded covalently to the MM region. In the QM calculations, these atoms are modeled by hydrogen atoms such that the QM region is truncated properly [3, 4]. DFT calculations with the resolution-of-the-identity (RI) approximation for the Coulomb terms [28], the Perdew-Becke-Ernzerhof exchange correlation functional [29] and a medium sized basis set (6-31(+)G*) are used to describe the QM region. Calculations are performed with the program Turbomole [30] and the effect from the surrounding MM atoms is treated as an external field in these calculations.

The above QM setup turns out not to be fully adequate for the description of the transition state. The free energy barrier can, however, easily be extrapolated to higher accuracy because energy fluctuations in a single simulation are fairly independent of the basis set and exchange-correlation functional chosen, see also Fig. 3. Here, we extrapolate to the Becke-3-Lee-Yang-Parr (B3LYP) functional [31] and a larger basis set (6-311++G(2d,2p)). This setup consumes 80 times more CPU time per QM/MM calculation than the former setup [32] and, thus, the extrapolation scheme saves significant computer time. More details of the calculations will be discussed elsewhere [33].

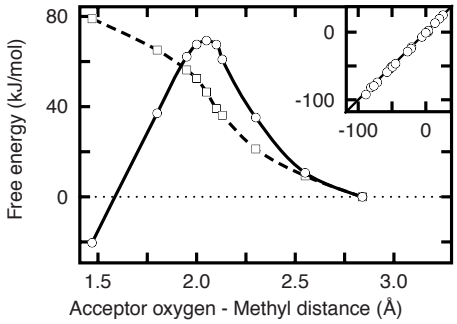

FIG. 3: Calculated barrier for the methyl transfer reaction catalyzed by catechol *O*-methyltransferase using the B3LYP hybrid functional and a large basis set. Forward reaction is from right to left. Dashed line indicates the contribution to the free energy from the MM-QM interaction. Error bars are not shown, but the statistical standard errors are less than 1 kJ/mol (for comparison, the height of a dot corresponds to ~ 4 kJ/mol). Inset: Correlation between QM/MM energy fluctuations calculated with the B3LYP functional and a large basis set (vertical axis) and the PBE functional and a medium sized basis set (horisontal axis).

The results are presented in Fig. 3 and show a barrier of 69 ± 1 kJ/mol (56 ± 1 without extrapolation). The re-

sults are well converged with statistical standard errors of less than 1 kJ/mol on the individual energy changes as well as for the activation energy [34]. This shows that it is feasible to calculate converged QM/MM free energy barriers based on the thermodynamic cycle in Fig. 1. In addition to the statistical uncertainty, there is a slight hysterisis of 1.6 kJ/mol for the transition state upon calculating the barrier for forward and backward reaction. This is a computational artifact, which is likely to disappear with denser sampling of the reaction pathway. The contribution to the free energy originating from interactions between the MM and QM region is included in Fig. 3 as well (dashed line) and shows how important a proper description of the environment is for an accurate calculation of the free energy barrier. If the environment described by MM atoms were omitted the reaction would occur almost spontanouesly. The importance of the environment has also been discussed in other computational studies [8, 17, 19].

The activation energy of the enzymatic reaction can be determined to be about 75 kJ/mol based on experimental results [35] in good agreement with our calculated one. The agreement may be caused by cancellation of errors. Both the force field and DFT methods have limited accuracy and the fixed reaction pathway spanned by relatively few degrees of freedom provides another potential source for errors. Moreover, we have ignored entropic changes associated with the relatively few degrees of freedom for the nuclei in the QM region [36]. Nonetheless, the current example demonstrates that it is feasible to apply the thermodynamic cycle in Fig. 1 for the calculation of converged high-level QM/MM free energy changes.

We are grateful for the help prodvided by Torben Rasmussen and Kristina Nilsson. We acknowledge and appreciate support given by The Swedish Research Council, Swedish National Infrastructure for Computing (SNIC) and LUNARC.

* Electronic address: Thomas.Rod@teokem.lu.se; URL: http://www.teokem.lu.se/~throd

[1] G. H. Jóhannesson, T. Bligaard, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Nørskov, Phys. Rev. Lett. **88**, 255506 (2002).
[2] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).
[3] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[4] U. Ryde, Curr. Opin. Chem. Biol. **7**, 136 (2003).
[5] S. Piana, P. Carloni, and M. Parrinello, J. Mol. Biol. **319**, 567 (2002).
[6] M. D. Peraro, L. I. Llarrull, U. Rothlisberger, A. J. Villa, and P. Carloni, J. Am. Chem. Soc. **126**, 12661 (2004).
[7] J. Bentzien, R. P. Muller, J. Florián, and A. Warshel, J. Phys. Chem. B **102**, 2293 (1998).
[8] M. Roca, S. Marti, J. Andrés, V. Moliner, I. Tuñon, J. Bertrán, and I. H. Williams, J. Am. Chem. Soc. **125**, 7726 (2003).
[9] S. J. Benkovic and S. Hammes-Schiffer, Science **301**, 1196 (2003).
[10] S. Muzammil, P. Ross, and E. Freire, Biochemistry **42**, 631 (2003).
[11] M. McGrath, I. Kawachi, A. Ascherio, G. A. Colditz, D. J. Hunter, and I. De Vivo, Am. J. Psychiatry **161**, 1703 (2004).
[12] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[13] D. L. Beveridge and F. M. DiCapua, Annu. Rev. Biophys. Biophys. Chem. **18**, 431 (1989).
[14] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).
[15] R. P. Muller and A. Warshel, J. Phys. Chem. B **99**, 17516 (1995).
[16] M. Štrajbl, G. Hong, and A. Warshel, J. Phys. Chem. B **106**, 13333 (2002).
[17] M. H. M. Olson, G. Hong, and A. Warshel, J. Am. Chem. Soc. **125**, 5025 (2003).
[18] T. Rasmussen, K. Nilsson, and U. Ryde, in preparation.
[19] B. Kuhn and P. A. Kollman, J. Am. Chem. Soc. **122**, 2586 (2000).
[20] E. Y. Lau and T. C. Bruice, J. Am. Chem. Soc. **122**, 7165 (2000).
[21] K. Kahn and T. C. Bruice, J. Am. Chem. Soc. **122**, 46 (2000).
[22] J. Vidgren, L. A. Svensson, and A. Liljas, Nature **368**, 354 (1994).
[23] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comp. Chem. **4**, 187 (1983).
[24] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, J. Phys. Chem. **97**, 10269 (1993).
[25] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[26] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, J. K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[27] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[28] K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, Theor. Chem. Acc. **97**, 119 (1997).
[29] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1995).
[30] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, Chem. Phys. Lett. **162**, 165 (1989).
[31] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).
[32] This difference is partly because the RI approximation cannot be used in connection with B3LYP in TURBOMOLE.
[33] T. H. Rod and U. Ryde, in preparation.
[34] Standard errors are calculated using a bootstrap procedure, Ref. [37]. Progression of errors give similar results.
[35] E. Schultz and E. Nissinen, Biochem. Phamacol. **38**, 3953 (1989).
[36] Entropy for the QM region can in principle be calculated from a normal mode analysis.
[37] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN: the art of scientific computing* (Cambridge University Press, Cambridge, 1992), 2nd ed.