# New Tools for the Accurate Study of Biomolecular Reactivity

The simulation of biochemical reactions has become a valuable counterpart to experimental investigations in rational drug design, catalysis, and the unravelling of natural processes. Such reactions pose the methodological challenge of a small reactive centre embedded in a large, inhomogeneous environment, e.g. proteins or DNA strands in aqueous solution.

Three levels of accuracy can be distinguished. (1) classical molecular mechanics force fields used in conjunction with molecular dynamics (MD) or Monte Carlo (MC) simulations. (2) Quantum mechanical methods to describe chemical reactions. (3) The QM/MM approach, which treats the reaction centre by quantum mechanics (QM) but incorporates the environment using classical molecular mechanics (MM). In this article we concentrate on some recent QM/MM enhancements to the programme package ChemShell which is developed by on-going collaboration between Computational Chemistry group at Daresbury, the group of Walter Thiel at MPI Mülheim and the group of Richard Catlow at the Royal Institution. The Mülheim group focuses on the biochemical tool development and it is this aspect of the collaboration, together with some recent applications work, that is summarised here.

Geometry optimisations provide minima and transition states on the potential energy surface. These are sufficient to describe the chemical reactivity of many small systems. The large, inhomogeneous, but comparably well-ordered environments of biochemical reactions, however, require the free energy, which includes the entropy, to be taken into account. We have developed a method to significantly improve the analysis of umbrella sampling simulations, biased MD simulations to calculate the free-energy change of chemical reactions.

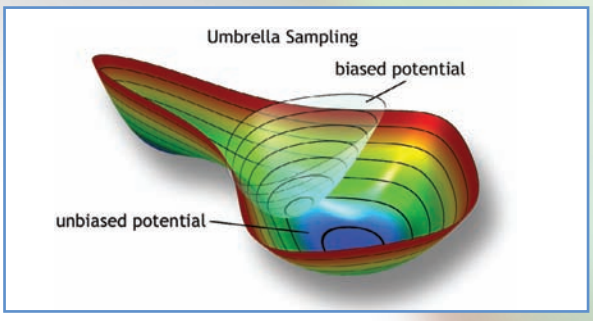


*Figure 1. In umbrella sampling the system evolves in a biased potential enabling efficient sampling of the transition state. Unbiasing using umbrella integration provides the free energy.*

## Umbrella Integration

Umbrella integration [1] is a method to analyse umbrella sampling simulations. It calculates the free energy by use of the mean force, which can directly be averaged over the steps, rather than weighting the probability density and finding the factors to combine the windows iteratively, as done in the standard weighted histogram analysis method (WHAM). Umbrella integration greatly reduces the noise inevitably present in MD simulations. The result is independent of other parameters of the analysis, such as the width of the bins for constructing the histograms in WHAM (see Figure 2).

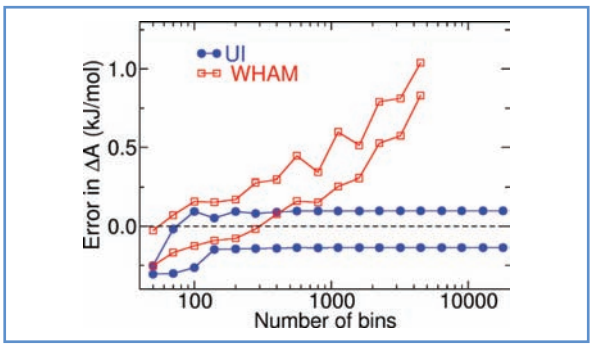


*Figure 2. Analysis of umbrella sampling simulations: umbrella integration (UI) converges, while no optimum number of bins can be given for the weighted histogram analysis method (WHAM).*

Umbrella integration moreover uses the MD trajectory to provide an estimate of the statistical error in the resulting free-energy difference [2]. This provides well-defined rules for how to optimally choose parameters for the umbrella sampling simulations, such as the strength of the bias, and the number of steps.

The drawback of umbrella sampling simulations, whether analysed with umbrella integration or with WHAM, is the necessity for extensive sampling. While this can easily be done for force fields, the computational costs prohibit sampling for quantum chemical or QM/MM simulations.

## QM/MM Free-Energy Perturbation

QM/MM free-energy perturbation allows the free energy change of the most important part of the system to be sampled at high levels of theory. The entropy and enthalpy changes of the QM/MM interactions and the MM part are sampled, while the entropy change of the QM part is neglected or estimated using the harmonic approximation. This allows the sampling to be confined to the MM part and classical MD to be used. The overall computational effort is dominated by geometry optimisations preceding the MD sampling. The QM/MM free-energy perturbation scheme has been implemented into ChemShell and demonstrated in a study of the catalytic bio-degeneration of aromatic compounds by p-hydroxybenzoate hydroxylase (PHBH). The QM/MM setup, containing about 22,700 MM atoms and 49 QM atoms, and the transition state, are depicted in Figure 3.

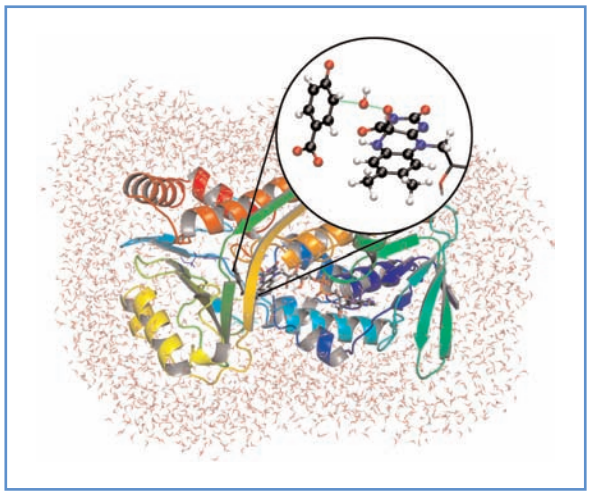

*Figure 3. The protein PHBH was simulated solvated in an 11Å-shell of water. The insert shows the transition state: QM atoms as ball-and-stick model, MM atoms as sticks.*

## QM/MM Studies on Metallo-Proteins

The ChemShell software has been used in extensive QM/MM studies of cytochrome P450 and related metallo-enzymes. Cytochrome P450 enzymes catalyze a great variety of stereospecific and regioselective oxygen insertion processes, which are of vital importance in biosystems for detoxification and biosynthesis. Given the biological importance of these enzymes and the fact that they are capable of activating inert C-H bonds under ambient conditions, the underlying reaction mechanisms have been intensely studied. The recent QM/MM investigations with ChemShell have addressed all relevant intermediates of the catalytic cycle, in particular the reactive Compound I species [4-6], and most of the key transformations in cytochrome P450cam. The QM/MM study of the mechanism of C-H hydroxylation supports a two-state two-step rebound mechanism in the doublet and quartet state [7], catalyzed by a single water molecule that is hydrogen-bonded to the oxo atom [8]. The QM/MM calculations on the formation of Compound I reveal a novel mechanism [9]. This work shows the great potential of QM/MM calculations for the atomistic modelling of enzymes in general, and for mechanistic studies of enzyme reactivity in particular. Such calculations provide unprecedented insight into the way enzymes work.

## Outlook

A range of approaches is necessary to tackle biochemical processes at high accuracy with reasonable computational costs. An additional dimension is added by calculating the free energy rather than the potential energy, which is achieved by sampling the phase space. These methods pave the way for studies with high predictive power complementing experimental data. Not only will they lead to a better understanding of biological processes, but they may lend guidance to the development of fine chemicals, biomimetic catalysts, and novel drugs.

*J. Kästner, P. Sherwood, CSED, Daresbury Laboratory;*
*H. M. Senn, S. Thiel, N. Otte, W. Thiel, MPI Mülheim, Germany.*

**References:**

*[1] J. Kästner, W. Thiel, J. Chem. Phys. 123, 144104 (2005).*

*[2] J. Kästner, W. Thiel, J. Chem. Phys. 124, 234106 (2006)*

*[3] J. Kästner, H. M. Senn, S. Thiel, N. Otte, W. Thiel, J. Chem. Theory Comput. 2, 452 (2006).*

*[4] J. C. Schöneboom, H. Lin, N. Reuter, W. Thiel, S. Cohen, F. Ogliaro and S. Shaik, J. Am. Chem. Soc. 124, 8142-8151 (2002).*

*[5] E. Derat, S. Cohen, S. Shaik, A. Altun and W. Thiel, J. Am. Chem. Soc. 127, 13611-13621 (2005).*

*[6] J. C. Schöneboom, F. Neese and W. Thiel, J. Am. Chem. Soc. 127, 5840-5853 (2005).*

*[7] J. C. Schöneboom, S. Cohen, H. Lin, S. Shaik and W. Thiel, J. Am. Chem. Soc. 126, 4017-4034 (2004).*

*[8] A. Altun, V. Guallar, R. A. Friesner, S. Shaik, and W. Thiel, J. Am. Chem. Soc. 128, 3924-3925 (2006).*

*[9] J. Zheng, D. Wang, W. Thiel, and S. Shaik, J. Am. Chem. Soc., online.*