



Mixed QM/MM Calculations in Biological Systems

G. M. Seabra, J. Torras, R. C. Walker, A. Roitberg

published in

*From Computational Biophysics to Systems Biology (CBSB07),
Proceedings of the NIC Workshop 2007,*
Ulrich H. E. Hansmann, Jan Meinke, Sandipan Mohanty,
Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. 36, ISBN 978-3-9810843-2-0, pp. 37-43, 2007.

© 2007 by John von Neumann Institute for Computing
Permission to make digital or hard copies of portions of this work for
personal or classroom use is granted provided that the copies are not
made or distributed for profit or commercial advantage and that copies
bear this notice and the full citation on the first page. To copy otherwise
requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume36>

Mixed QM/MM Calculations in Biological Systems

Gustavo de M. Seabra¹, Juan Torras², Ross C. Walker³, and Adrian Roitberg¹

¹ Quantum Theory Project, University of Florida
P.O. Box 118435, Gainesville, FL 32611-8435, USA
E-mail: roitberg@ufl.edu

² Departament d'Enginyeria Química, EUETII, Universitat Politècnica de Catalunya
Pça. Rei 15, 08700-Igualada, SPAIN

³ San Diego Supercomputer Center, University of California San Diego
9500 Gilman Drive, La Jolla, CA, 92093-0505, USA

The computational study of a variety of important processes, such as processes involving bond breaking and forming and electron reorganization, requires the application of quantum mechanical methods. In biological systems, this situation is further complicated because the influence of the environment must be taken into account. In these situations, hybrid quantum mechanics / molecular mechanics methods can be used. We describe here two new QM/MM implementations based on the Amber Molecular Dynamics packages that make it possible to use of a variety of quantum mechanical methods with any of the techniques available in Amber. Results are provided for the application of these implementations to two different systems.

1 Introduction

Due to the large system size, computational studies of biologically relevant systems such as proteins, nucleic acids and carbohydrates need to resort to approximated Molecular Mechanics (MM) methods which apply parameterized force fields to describe molecular properties such as bond lengths, angles, dihedrals, electrostatic and van der Waals forces. The use of such parameterized methods greatly reduces the computational complexity, allowing the study of processes including ligand binding, enzyme reaction mechanisms, protein folding, refolding, and denaturation, providing invaluable help in the analysis of complex experimental data and structures.

Despite continuous efforts to develop more reliable force fields for use in MM calculations, including the use of QM calculations and genetic algorithms in the parameterization of the force field, classical mechanics methods lack the ability to treat fundamentally quantum processes, such as bond breaking and forming and charge fluctuations as a function of geometry, or to describe parts of the potential energy surface far from equilibrium. In some cases, although computationally expensive, it is possible to treat a model system purely by QM methods, but the effect of the environment must be either neglected or simulated by a continuum dielectric approximation. An alternative that allows the explicit inclusion of environment effects while treating the most relevant part of the system with full quantum mechanics was first explored by Warshel and Levitt as early as 1976,¹ and is the use of hybrid quantum mechanics / molecular mechanics (QM/MM) calculations whereby a subsection of the system is treated by QM methods, the remainder (environment) is treated by standard molecular mechanics (MM) methods, and a coupling potential is used to connect the two regions.

Recent work in our group has focused in the implementation and application of hybrid QM/MM techniques for the study of biologically relevant systems. In particular, discussed here are two recently developed interfaces: the new native, (semi-empirical) QM/MM support available in Amber 9,² and the integration of the Amber MD program with a QM program through the PUPIL interface. These interfaces offer the advantage of blending seamlessly with the Amber program, allowing the application of any of the advanced sampling methods available in Amber to QM/MM problems.

2 QM/MM

In a hybrid QM/MM calculation, the system is partitioned into two regions: A QM region, typically consisting of a relatively small number of atoms relevant for the specific process being studied, and a MM region with all the remaining atoms. The total Hamiltonian (\hat{H}) for such a system is written as:

$$\hat{H} = \hat{H}^{QM} + \hat{H}^{MM} + \hat{H}^{QM/MM}, \quad (1)$$

where \hat{H}^{QM} and \hat{H}^{MM} are the Hamiltonians for the QM and MM parts of the system, and are calculated using either the QM method chosen or the usual force field equations, respectively. The remaining term, $\hat{H}^{QM/MM}$, describes the interaction between the QM and MM parts and typically contains terms for electrostatic, van der Waals and bonded interactions across the region boundaries:

$$\hat{H}^{QM/MM} = \hat{H}_{vdW}^{QM/MM} + \hat{H}_{elect}^{QM/MM} + \hat{H}_{bonds}^{QM/MM}. \quad (2)$$

In the approach used for both Amber QM/MM interfaces presented here, the van der Waals ($\hat{H}_{vdW}^{QM/MM}$) is treated as usual by the MM program, using the standard 12-6 Lennard-Jones parameters from the force field in use. It has been shown that this choice does not introduce significant errors in the calculation.³ Also, both interfaces described here use a link atom scheme if bonds are broken across the QM–MM boundary. Finally, the remaining term ($\hat{H}_{elect}^{QM/MM}$), describing the electrostatic interaction between the MM and QM zones, depends on the specific interface.

3 Semi-Empirical QM/MM in Amber

One of the most used programs for MD simulations is the *sander* program, part of the Amber suite. The *sander* interface for QM/MM has been recently rewritten, placing strong emphasis on speed and accuracy, allowing the simulation of systems with a reasonably sized quantum mechanical region (around 300 atoms) for long (nanoseconds) timescales. It now natively includes a number of semi-empirical Hamiltonians for QM/MM simulations including MNDO,^{4,5} AM1,⁶ PM3,^{7,8} PDDG/PM3⁹ and SCC-DFTB.¹⁰ Our group has been involved in the implementation of the SCC-DFTB method.

The Self-Consistent-Charge Density-Functional Tight Binding (SCC-DFTB) is an approximate method based on Density Functional Theory (DFT) from a second order expansion of the DFT total energy in the charge density fluctuations, and has been described in detail elsewhere.^{11,10} It has been successfully applied to the study of a variety of systems, and has been shown to yield results comparable in accuracy to *ab-initio* MP2 calculations

with large basis sets.¹² (For a recent review on applications of SCC-DFTB to biological systems, the reader is referred to Ref.[13].)

The integration of SCC-DFTB with Amber 9 has been described recently.¹⁴ In this implementation, the electrostatic interaction energy between the QM and MM regions ($\hat{H}_{elect}^{QM/MM}$ in Equation 2) is calculated as a Coulomb interaction between the atomic Mulliken charges calculated from the SCC-DFTB (q_α) and the classical RESP charges on the MM atoms (Q_A), as in Equation 3.

$$E_{elect}^{QM/MM} = \sum_{\alpha}^{QM} \sum_A^{MM} \frac{Q_A q_\alpha}{r_{\alpha A}}. \quad (3)$$

Showing how the QM/MM implementation can directly tap into the methods in Amber, the free energy surface (FES) for the capped alanine dipeptide (ACE-ALA-NME, AD) in vacuum was generated using Replica Exchange Molecular Dynamics,¹⁵ with the AD treated by SCC-DFTB. 6 replicas were used, at temperatures of 161.2K, 219.9K, 300.0K, 419.3K, 558.4K and 761.8K. Exchanges were attempted 10,000 times, with 0.5ps between attempts, and a time step of 1fs with SHAKE was used. The Langevin thermostat with a

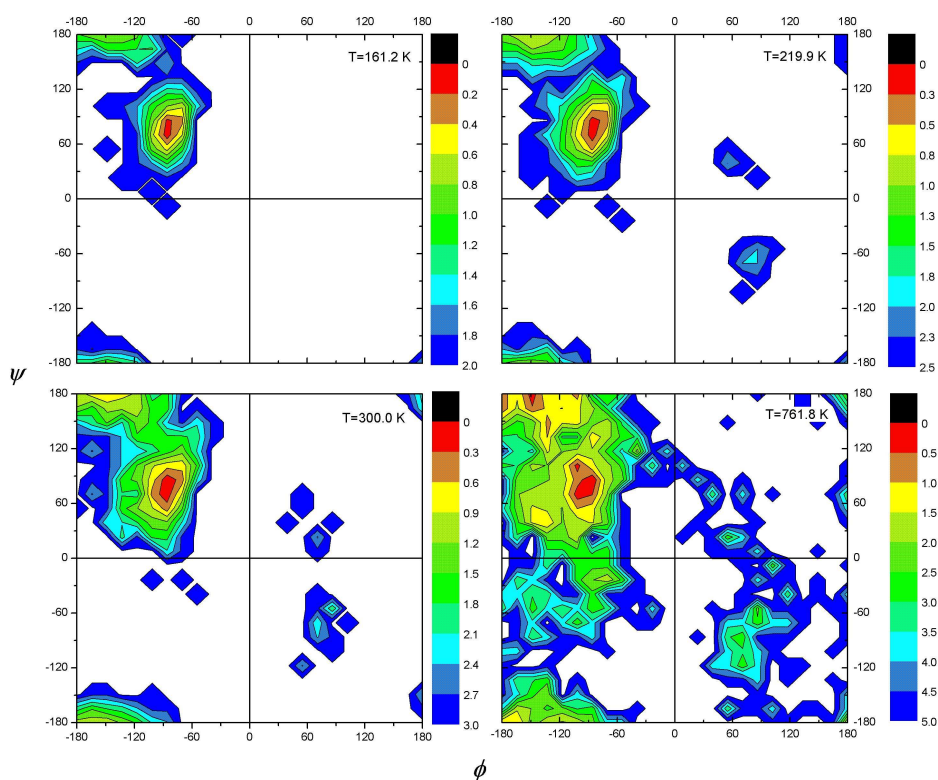


Figure 1. Free Energy Surfaces for alanine dipeptide obtained from the SCC-DFTB/REMD calculation.

collision frequency of 2.0ps^{-1} was used to regulate the temperatures.

Figure 1 shows the FES for four different temperatures, as calculated from $G = -RT \ln(P)$, where G is the Gibbs free energy, R is the gas constant and T is the temperature, and P is the (normalized) probability P of finding the AD in a conformation at a particular region in (ϕ, ψ) -space from the MD trajectories. The surfaces show a clear minimum around ϕ/ψ values of $(-83^\circ, 76^\circ)$, which corresponds to the known $C7_{eq}$ minimum for AD in vacuum, and other structures are populated as temperature increases. The relative energies compare well with calculated *ab-initio* values calculated previously.¹⁶

4 Amber-Gaussian QM/MM Through the PUPIL Interface

The interface described in the previous session has the advantage of being convenient and easy to use, especially for users experienced with the Amber MD program. At the current stage of development, however, the QM region is fixed throughout the calculation, and limited to semi-empirical Hamiltonians.

The PUPIL (Program for User Program Interfacing and Linking) is a free, open source package created to facilitate the interfacing of *arbitrary* MD and QM programs.^{17,18} Currently, interfaces with DLPOLY, MNDO97 and Siesta have been developed. We have

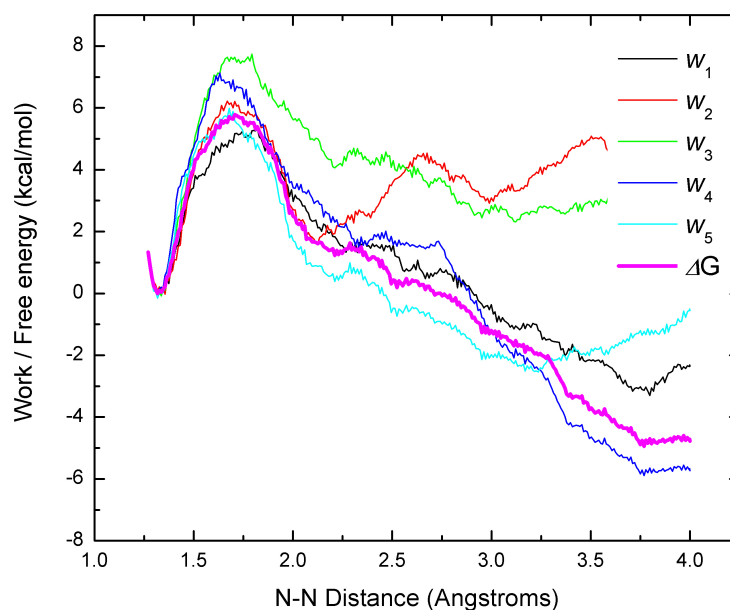


Figure 2. Work of breaking the N-N bond in Angeli's salt for 5 different simulations, and the free energy obtained by the Jarzynski equation.

recently included interfaces with *sander* and Gaussian03 to the PUPIL system. In PUPIL, the equations of motion are propagated as usually by the MD program and the forces over the QM region are provided by the QM program, while the PUPIL interface manages execution of the two programs, as well as the required information flow. The influence of the MM atoms on the QM zone is taken into account by electronic embedding, where the MM atoms are considered as point charges fixed at the positions of the MM atoms, and with the charge values from the force field parameters in use.

In the Gaussian03 interface, the force contribution from the quantum atoms to the total force acting upon the classical atoms (F_i^{QM}) is obtained by dividing the electronic density into a grid, and calculating the interaction of the MM charge with each point in the grid:

$$F_i^{QM} = \sum_j^{cube} \mathbf{r}_{ij} \frac{q_{PC_i} dq_j}{|r_{ij}|^3}, \quad (4)$$

where

$$dq_j = \rho_j dx dy dz. \quad (5)$$

As an example, the free energy change associated to a specific process can be calculated using the Jarzynski relationship:¹⁹⁻²¹

$$e^{-\beta\Delta G(d)} = \left\langle e^{-\beta W(d)} \right\rangle_{d=d_0}, \quad (6)$$

where the brackets indicate an average taken over a large number of independent realizations of the process (in this case, molecular dynamics trajectories) all starting at different points belonging to the same equilibrated ensemble.

Figure 2 shows work results five trajectories breaking the N-N bond in the monoprotonated Angeli's salt (O2N-NHO⁻) in explicit water, together with the free energy calculated using Equation 6. The Angeli's salt was treated quantum mechanically by Gaussian03 at the UB3LYP/6-311+G(d) level, and put in a large box of TIP3P water molecules. The free energy barrier for breaking the N-N bond is calculated to be 5.84 kcal/mol, which is in good agreement with previous estimations using B3LYP with the same basis set and implicit solvation.²²

5 Conclusions

This communication describes and exemplifies two new QM/MM implementations. Both have the advantage of being based on the Amber molecular dynamics package, allowing their application in tandem with any of the advanced sampling methods available in Amber. Results of calculations using both implementations are provided.

Acknowledgments

The authors want to acknowledge Erik Deumens and S.B. Trickey for their part in the initial development of the PUPIL package. Computer resources were provided by the Large Allocations Resource Committee through TeraGrid grant TG-MCA05S010 to AER, and

the TeraGrid grant TG-CHE060072T to GS and JTC, the High-Performance Computing Center at the University of Florida and a local cluster built with a generous IBM-SUR grant. RCW acknowledges NIH Grant GM57513 for funding this work. RCW and AER would like to thank the San Diego Supercomputer Center for their continued support of this project and future developments in Amber 10 through their Strategic Applications Collaboration program.

References

1. A. Warshel and M. Levitt, *Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme*, Journal of Molecular Biology, **103**, no. 2, 227–249, 1976.
2. “Amber 9”, <http://amber.scripps.edu>.
3. Demian Riccardi, Guohui Li, and Qiang Cui, *Importance of van der Waals Interactions in QM/MM Simulations*, Journal of Physical Chemistry B, **108**, 6467–6478, 2004.
4. Michael J. S. Dewar and Walter Thiel, *Ground states of molecules. 38. The MNDO method. Approximations and parameters*, J. Am. Chem. Soc., **99**, no. 15, 4899–4907, 1977.
5. Michael J. S. Dewar and Walter Thiel, *Ground states of molecules. 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen*, J. Am. Chem. Soc., **99**, no. 15, 4907–4917, 1977.
6. Michael J. S. Dewar, Eve G. Zoebisch, Eamonn F. Healy, and James J. P. Stewart, *Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model*, J. Am. Chem. Soc., **107**, no. 13, 3902–3909, 1985.
7. James J. P. Stewart, *Optimization of parameters for semiempirical methods I. Method*, Journal of Computational Chemistry, **10**, no. 2, 209–220, 1989.
8. James J. P. Stewart, *Optimization of parameters for semiempirical methods II. Applications*, Journal of Computational Chemistry, **10**, no. 2, 221–264, 1989.
9. Matthew P. Repasky, Jayaraman Chandrasekhar, and William L. Jorgensen, *PDDG/PM3 and PDDG/MNDO: Improved semiempirical methods*, Journal of Computational Chemistry, **23**, no. 16, 1601–1622, 2002.
10. M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and Gotthard Seifert, *Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties*, Physical Review B: Condensed Matter, **58**, no. 11, 7260–7268, 1998.
11. Thomas Frauenheim, D. Porezag, Marcus Elstner, G. Jungnickel, J. Elsner, M. Haugk, A. Sieck, and Gotthard Seifert, *An ab initio two-center tight binding approach to simulations of complex materials properties*, Mat. Res. Soc. Symp. Proc., **491**, 91–104, 1998.
12. H. G. Bohr, K. J. Jalkanen, M. Elstner, K. Frimand, and S. Suhai, *A comparative study of MP2, B3LYP, RHF and SCC-DFTB force fields in predicting the vibrational spectra of N-acetyl-L-alanine-N^γ-methyl amide: VA and VCD spectra*, Chemical Physics, **246**, no. 1-3, 13–36, 1999.

13. M. Elstner, *The SCC-DFTB method and its application to biological systems*, Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta), **116**, no. 1-3, 316–325, 2006.
14. Gustavo M. Seabra, Ross C. Walker, Marcus Elstner, David A. Case, and Adrian E. Roitberg, *Implementation of the SCC-DFTB Method for Hybrid QM/MM Simulations within the Amber Molecular Dynamics Package*, Journal of Physical Chemistry A, **111**, no. 26, 5655–5664, 2007.
15. Ulrich H. E. Hansmann, *Parallel tempering algorithm for conformational studies of biological molecules*, Chemical Physics Letters, **281**, no. 1-3, 140–150, 1997.
16. Rubicelia Vargas, Jorge Garza, Benjamin P. Hay, and David A. Dixon, *Conformational Study of the Alanine Dipeptide at the MP2 and DFT Levels*, Journal of Physical Chemistry A, **106**, no. 13, 3213–3218, 2002.
17. Juan Torras, E. Deumens, and S. B. Trickey, *Software Integration in Multi-scale Simulations: the PUPIL System*, Journal of Computer-Aided Materials Design, **13**, no. 1, 201–212, 2006.
18. Juan Torras, Yao He, Chao Cao, Krishna Muralidharan, E. Deumens, H. P. Cheng, and S. B. Trickey, *PUPIL: A systematic approach to software integration in multi-scale simulations*, Computer Physics Communications, **177**, no. 3, 265–279, 2007.
19. C. Jarzynski, *Nonequilibrium equality for free energy differences*, Physical Review Letters, **78**, no. 14, 2690–2693, 1997.
20. A. Crespo, M. A. Marti, D. A. Estrin, and A. E. Roitberg, *Multiple-Steering QM-MM Calculation of the Free Energy Profile in Chorismate Mutase*, J. Am. Chem. Soc., **127**, no. 19, 6940–6941, 2005.
21. Hui Xiong, Alejandro Crespo, Marcelo Marti, Dario Estrin, and Adrian Roitberg, *Free Energy Calculations with Non-Equilibrium Methods: Applications of the Jarzynski Relationship*, Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta), **116**, no. 1-3, 338–346, 2006.
22. A.S. Dutton, J.M. Fukuto, and K.N. Houk, *Mechanisms of HNO and NO Production from Angeli's Salt: Density Functional and CBS-QB3 Theory Predictions*, J. Am. Chem. Soc., **126**, no. 12, 3795–3800, 2004.