ChemShell—a modular software package for QM/MM simulations

Sebastian Metz,1 Johannes Kästner,2 Alexey A. Sokol,3 Thomas W. Keal1* and Paul Sherwood1

ChemShell is a modular computational chemistry package with a particular focus on hybrid quantum mechanical/molecular mechanical (QM/MM) simulations. A core set of chemical data handling modules and scripted interfaces to a large number of quantum chemistry and molecular modeling packages underpin a flexible QM/MM scheme. ChemShell has been used in the study of small molecules, molecular crystals, biological macromolecules such as enzymes, framework materials including zeolites, ionic solids, and surfaces. We outline the range of QM/MM coupling schemes and supporting functions for system setup, geometry optimization, and transition-state location (including those from the open-source DL-FIND optimization library). We discuss recently implemented features allowing a more efficient treatment of long range electrostatic interactions, X-ray based quantum refinement of crystal structures, free energy methods, and excited-state calculations. ChemShell has been ported to a range of parallel computers and we describe a number of options including parallel execution based on the message-passing capabilities of the interfaced packages and task-farming for applications in which a number of individual QM, MM, or QM/MM calculations can performed simultaneously. We exemplify each of the features by brief reference to published applications. © 2013 The Authors. WIREs Computational Molecular Science published by John Wiley & Sons, Ltd.

INTRODUCTION

Combined quantum mechanical (QM) and molecular mechanical (MM) modeling is a widely used technique to perform atomistic simulations of large systems as encountered in materials modeling and biochemistry. In a QM/MM calculation, the QM description is restricted to the central region where information about local electronic properties is required, while the environment is modeled with a much cheaper classical molecular mechanics force field. In this way, the simulation becomes tractable without compromising significantly on the accuracy of the final result, whereas a QM simulation of the whole system may have an unnecessarily high or even prohibitive computational cost.

In order to perform a QM/MM calculation, the software package must be able to run both QM and MM calculations and be able to combine the results in a meaningful way, including an appropriate treatment of both through space (nonbonded) interactions and covalent bonds between the QM and MM regions. Historically, most QM and MM programs have been developed as separate packages. More recently, a number of QM packages have had MM routines added to them, and vice versa. An alternative approach, which we believe offers greater flexibility, is to combine the output of separate fully featured QM and MM packages using a software wrapper. The advantage of this modular approach is that the user may in principle mix and match QM and
MM codes that implement the features of interest to them. The ChemShell computational chemistry environment is an implementation of this approach.\textsuperscript{1} ChemShell performs a QM/MM calculation by decomposing the system to be modeled into QM and MM subsystems, interfacing to external programs to perform the QM and MM calculations and then combining the resulting energies and gradients to obtain a QM/MM energy and gradient. ChemShell can also be used to drive higher level chemical simulation tasks such as geometry optimization or molecular dynamics (MD), using QM, MM, or QM/MM Hamiltonians.

In this article, we describe the capabilities of version 3.5 of ChemShell, released in May 2012, except where features are noted as under development. Original references for methodological developments are not in general given, but may be found in the implementation and application works we cite. The selected applications presented here are chosen to illustrate the described features and are in no way exhaustive. A more detailed introduction to the different flavors of QM/MM and applications of the method (not only with ChemShell) can be found elsewhere.\textsuperscript{2,3}

**THE CHEMSSHLL FRAMEWORK**

ChemShell is based on the Tcl scripting language and as such it can be run interactively or presented with a script file describing the sequence of calculations required. All the expected features of a scripting language (string handling, conditional statements, loops, etc.) are augmented with additional commands to perform molecular modeling operations. Tasks such as defining a molecular structure or performing a geometry optimization are initiated by simple commands, forming with three QM/MM coupling schemes. The primary strength lies in combined QM/MM calculations. It is helpful to divide QM/MM schemes by the way the total energy is defined. We use the term ‘additive’ for schemes combining a QM calculation on the inner region, an MM calculation on the outer region, and the QM/MM coupling terms. Most calculations performed with ChemShell are of this type. Additive QM/MM calculations may require modifications to the force field terms applied (e.g., in the junction region) and this restricts the choice of MM packages. Alternatively, a QM calculation on the inner region can be combined with a MM calculation on the entire system with a third MM calculation on the inner region being performed to subtract out the terms that are being replaced by the QM Hamiltonian. Such ‘subtractive’ QM/MM calculations can be performed in ChemShell using any QM and MM codes, or indeed can be used to combine two different QM Hamiltonians in an analogous way.

Additive QM/MM calculations can be performed with three QM/MM coupling schemes. The most basic is termed ‘mechanical embedding’, where the QM calculation is not coupled electrostatically to the MM environment. The second level is ‘electrostatic embedding’, where classical partial charges from the MM description enter the QM Hamiltonian, in effect polarizing the QM region. The third level is ‘polarized MM embedding’, where the MM region is in turn polarized by the QM region. Polarized MM embedding may place additional restrictions on the choice of QM and MM codes (see Table 1). The level of embedding used depends on how accurately the electrostatic interaction between the QM and MM

When a calculation using an interfaced package is performed, ChemShell creates the relevant input files based on the control parameters presented in the user’s script. These parameters (which define such choices as basis set, level of theory, force field, etc.) are defined in a consistent way between similar types of external packages, allowing the user to switch between alternative packages in a relatively straightforward manner. The list of QM and MM codes that can be interfaced to ChemShell is shown in Table 1. Each code has a corresponding interface module in ChemShell, mostly written in Tcl to allow for extensions to support additional features in the external codes as they are developed.

**ESTABLISHED QM/MM METHODOLOGY**

ChemShell can be (and is often) used to script standalone QM or MM calculations. However, its primary strength lies in combined QM/MM calculations. Additive QM/MM calculations can be performed with three QM/MM coupling schemes. The most basic is termed ‘mechanical embedding’, where the QM calculation is not coupled electrostatically to the MM environment. The second level is ‘electrostatic embedding’, where classical partial charges from the MM description enter the QM Hamiltonian, in effect polarizing the QM region. The third level is ‘polarized MM embedding’, where the MM region is in turn polarized by the QM region. Polarized MM embedding may place additional restrictions on the choice of QM and MM codes (see Table 1). The level of embedding used depends on how accurately the electrostatic interaction between the QM and MM

...
TABLE 1 | External QM and MM Codes that can be Interfaced to ChemShell

| QM/MM features | HPC features |
|----------------|-------------|
| Electrostatic embedding | Directly linked interface |
| Shell model embedded cluster calculations | Task-farming parallelization (directly linked) |

| QM codes | Electrostatic embedding | Shell model embedded cluster calculations | Multiple state interface | Directly linked interface | Task-farming parallelization (directly linked) |
|----------|------------------------|------------------------------------------|-------------------------|--------------------------|-------------------------------------------|
| GAMESS-UK | X                      | X                                        |                         | X                        | X                                         |
| NWChem | X                      | X                                        |                         | X                        |                                           |
| FHI-AIMS¹ | X                      | X                                        |                         |                          |                                           |
| DALTON | X                      |                                           |                         |                          |                                           |
| LS-DALTON¹ | X                      |                                           |                         |                          |                                           |
| DMol3 | X                      |                                           |                         |                          |                                           |
| TURBOMOLE | X                      |                                           |                         |                          |                                           |
| ORCA | X                      |                                           |                         |                          |                                           |
| GAMES (US) |                       |                                           |                         |                          |                                           |
| Q-Chem | X                      |                                           |                         |                          |                                           |
| Gaussian² | X                      | X                                        |                         |                          |                                           |
| Molpro | X                      |                                           |                         |                          |                                           |
| MNDO | X                      |                                           |                         |                          |                                           |
| MOPAC | X                      |                                           |                         |                          |                                           |
| MM codes |                       |                                           |                         |                          |                                           |
| DL_POLY 2² | X                      |                                           |                         |                          |                                           |
| GULP | X                      |                                           |                         |                          | X                                         |
| GROMOS | X³                      |                                           |                         |                          | X                                         |
| CHARMM | X³                      |                                           |                         |                          | X                                         |

¹Scheduled for release in ChemShell 3.6.
²Not maintained by Daresbury Laboratory due to licensing restrictions.
³Internal module with integrated support for CHARMM and AMBER force fields.
⁴Possible but there is no automatic deletion of duplicate FF terms and so the user is responsible for providing a suitable topology.

region should be described, and whether the preferred MM description incorporates polarization.

There are two approaches to handling the QM/MM boundary: covalent and ionic. The covalent approach is the obvious choice whenever covalent bonds connect the QM and MM regions, as is usually the case in studies of enzymes and inorganic networks with covalent character, such as zeolites. ChemShell implements a ‘link atom’ approach in which the broken bonds are terminated by adding hydrogen atoms to the QM region while force field terms in the MM calculation are selectively deleted to ensure that no interactions are double counted. Further, the charges on the MM atoms closest to the QM region are shifted to neighboring atoms to avoid overpolarization of the QM system. Extra dipoles are then added to compensate for the dipoles created by the charge shifting scheme. Finally, the forces on the link atoms are projected onto real atoms after each gradient evaluation. For further details see Ref 1.

For some dense materials, such as metal oxides, an ionic approach (as described below) is more appropriate. Link atoms are not necessary as there are no covalent bonds to break. Instead atomic, large-core pseudopotentials are placed on the MM atoms immediately surrounding the QM region in order to localize electrons in the QM region, preventing them from spilling out into the MM region.⁴

Finally, if there are no covalent or ionic bonds that cross the QM and MM regions, no specific boundary handling method is required, as is the case in many solvation studies and molecular crystals⁵ (assuming that the QM/MM boundary does not divide any of the molecules).

Biomolecular Calculations

The recommended MM interface for biomolecular QM/MM calculations is the internal DL_POLY module, based on the DL_POLY 2 (now ‘DL_POLY Classic’) code developed by W. Smith.⁶ Routines are provided to convert force field parameter and topology files from formats such as CHARMM and AMBER into an internal format suitable for calculations with DL_POLY. A common workflow is to set up and equilibrate the system initially in CHARMM or AMBER before importing the coordinates and force field into ChemShell to perform a QM/MM optimization.

Most biochemical simulations to date using ChemShell have used the electrostatic level of
embedding, where the QM region is polarized by partial charges as defined in the standard biomolecular force fields (illustrated in Figure 1). This approach is exemplified by an extensive series of studies of the cytochrome P450 superfamily. Although these enzymes share the same active species, they differ in the immediate protein environment. The QM/MM methodology is an ideal approach to explore the differences as it allows these environmental variations to be compared at little extra cost. After over 10 years of intensive research, QM/MM results have provided descriptions of the various species present in the catalytic cycle and an understanding of their molecular and chemical properties. Numerous mechanisms by which the native enzymes, and mutants thereof, catalyze the oxygenation reactions have been used to rationalize experimental findings and predictive models of mechanisms and reactivity have been established. One of the key intermediates of the catalytic cycle (Cpd I) was predicted in QM/MM studies a long time before its experimental discovery and the simulations and predictions about the selenocysteine mutant also preceded experimental preparation. A thorough review of the simulated results from the first QM/MM study in 2002 up to 2010 is available.7 However, investigations on similar systems do and will continue.8–10

In the area of biomolecular modeling, ChemShell has been applied to a broad variety of enzymes to investigate their reaction mechanisms, for example, studies on the C–F bond-forming step in fluorinase, catalyzing the reaction between S-adenosyl-l-methionine and fluoride ions to form 5′-fluoro-5′-deoxyadenosine and L-methionine,11 and the metabolic oxidation of xanthine to uric acid performed by xanthine oxidase.12 In addition to obtaining structural features of intermediates, QM/MM calculations can be used to predict spectroscopic properties, like experimentally easily accessible 1H or 13C NMR shifts,13 NMR shifts of transition metal centers like vanadium,14,15 UV–vis, Raman, and IR data16 or EPR tensors.17

**Solid-State Systems**

ChemShell implements a finite cluster embedding model for QM/MM solid state calculations of bulk defects and surface reactions, as illustrated in Figure 2. In this approach, a spherical or hemispherical cluster is cut from a periodic MM model of the material. Point charges are then placed in a shell around the cluster and their charges set to reproduce as closely as possible the missing classical periodic electrostatic environment. A QM region is then defined in the center of the cluster and additional reactants added as appropriate. There are several advantages to using a finite representation of the solid in QM/MM calculations: spurious periodic interactions between defects are avoided; molecular QM codes, which lack support for periodic boundary conditions can still be used; and charged species are straightforward to model.

As of version 3.4, ChemShell contains integrated setup routines for extended covalent, ionic, and molecular crystal systems. These have a common interface and differ mainly in how the outer boundary
of the MM cluster is handled: for example, the molecular crystal routines ensure that only whole molecules are included in the cluster. The type of system also determines how the QM/MM boundary should be handled. For molecular crystals and semi-ionic solids such as zeolites, the link atom approach may be used. The DL_POLY interface is recommended for these calculations. Solid-state classical force field parameters may be defined explicitly using this interface. For calculations on zeolites, the MM charges may be adjusted automatically to account for the usually nonzero charge of the QM region (in addition to the normal charge shift procedure).

In order to model molecular crystals accurately, it is necessary to obtain appropriate point charges for the MM region. A protocol has been developed to obtain point charge values based on DFT calculations updated self-consistently in the field of point charges, which is available as a ChemShell script.

The ionic approach to embedding, used in materials such as metal oxides, may be employed using either electrostatic embedding (using a rigid-ion force field) or polarized MM embedding using a shell model potential. Because of the strong electrostatic interactions in an ionic material it is often necessary to use a shell model potential to obtain an accurate QM/MM result. The shell model splits the charges on classical atoms into core and shell components, connected by a spring force. The shell component represents the valence electrons that can be polarized by the QM region. In a QM/MM calculation, the shell positions are relaxed self-consistently by multiple alternating QM and MM calculations at every step. The GULP MM package is used to perform shell model calculations in ChemShell, and QM/MM shell model force fields may be defined from a standard GULP format force field using an automated script.

Applications of the ionic embedded cluster model include the study of point defects in zinc oxide and catalysis on zinc oxide surfaces (reviewed elsewhere) and the formation of water on interstellar olivine grains.

**RECENT DEVELOPMENTS**

**Geometry Optimization with DL-FIND**

Characterizing chemical reaction paths using geometry optimization algorithms is one of the most fundamental tasks in computational chemistry. A number of optimization modules are available in ChemShell, but the recommended optimizer is DL-FIND. DL-FIND offers a range of algorithms to explore potential energy surfaces and can be applied to the location of energy minima, transition states (including the dimer method), reaction path optimization (nudged elastic band), conical intersection search, and global optimization through stochastic search and a genetic algorithm. Several coordinate systems are available internally in DL-FIND, including the efficient hybrid delocalized coordinate system originally developed for the HDLCOpt optimizer.

Large-scale systems optimizations may be performed using the low memory BFGS algorithm. A microiterative minimization scheme for efficient QM/MM optimization has also been developed in ChemShell’s HDLCOpt optimizer and the method is currently being reimplemented in DL-FIND with extensions to transition state and reaction path optimization. In microiterative optimization, the MM
environment is relaxed after every QM step in the optimization, thereby reducing the number of expensive QM evaluations at the cost of a larger number of cheaper MM evaluations, giving a lowered overall computational expense.

DL-FIND has also recently been extended to perform reaction rate calculations. Reactant and transition state Hessian calculations may be used to calculate rates by harmonic transition-state theory, or instanton theory may be used to optimize the most likely quantum tunneling path at a given temperature.

DL-FIND is distributed with ChemShell but is also available as a standalone open source library with a well-defined interface for use with other codes. It has been interfaced to QM codes such as GAMESS-UK, CRYSTAL, and CASTEP as well as MM routines from DL_POLY and Tinker.

Treatment of Long-Range Electrostatic Interactions

Long-range electrostatic interactions within a macromolecule/solvent system may be modeled efficiently using one of two boundary potential methods implemented in ChemShell: the generalized solvent boundary potential (GSBP) and the solvated macromolecule boundary potential (SMBP). In these approaches, an inner atomistic region is defined surrounded by an outer region which is represented implicitly by the boundary potential. The GSBP requires modifications to the QM Hamiltonian and is only implemented in the MNDO semiempirical code, and is targeted at MD simulations where the high initial setup overhead is outweighed by the large number of subsequent calculations. By contrast the SMBP represents the boundary potential due to the outer parts of the investigated macromolecular system and bulk solvent solely by point charges calculated on-the-fly, and is therefore suited to single point calculations and geometry optimizations, providing an alternative to simply ‘chopping’ down the system by means of a distance criterion. The SMBP can be used with any QM code that supports electrostatic embedding.

The effect of the SMBP method has been analyzed by measuring the influence of its usage on the energetics of previously obtained reaction mechanisms, namely an intramolecular Claisen rearrangement in chorismate mutase and a hydroxylation reaction in p-hydroxybenzoate hydroxylase. Results show that there is a need for SMBP when the reaction under consideration is accompanied by a significant rearrangement of the charge distribution.

Recently, a polarized MM embedding scheme has been implemented in ChemShell based on the charge-on-spring/Drude oscillator approach, which modifies the MM force field to include terms that are polarized by the electrostatic potential of the QM region. This is conceptually very similar to the shell model described above for solid-state studies. It has further been combined with SMBP and GSBP leading to a fully polarizable three-layer method.

X-Ray Based Quantum Refinement of Crystal Structures

The QM/MM driver in ChemShell has been extended to include crystallographic X-ray data through an interface to the CNS (Crystallography & NMR System) package. Standard crystal structure refinement, especially for protein structures, is often carried out based on classical force fields. As the main part of proteins consist of amino acids for which thoroughly tested parameters are available, this is of little concern. The force field parameters for cofactors, substrates, inhibitors, metals centers, and so on are however much less reliable. Rather than optimizing the crystal structure based on a MM energy expression, the refinement can be performed in ChemShell using QM/MM information. The method has been applied to investigate the chromophore structure of the red fluorescent protein DsRed.M1 focusing on the cis and trans conformers of the acylimine chromophore, providing additional insight into the hydrogen bonding network within the protein.

MD and Free Energy Calculations

The MD module in ChemShell is based on routines from DL_POLY. Standard MD functionality such as NVE, NVT, and NPT ensembles, constraints using the SHAKE procedure, and rigid groups are available. An MD calculation may be performed with any code interfaced to ChemShell so that QM, MM, and QM/MM MD simulations are all possible. The QM/MM MD method has been outlined for the case of the hydroxylation reaction in the enzyme p-hydroxybenzoate hydroxylase using a semiempirical QM method, in which free energies of activation were obtained by thermodynamic integration.

Free energy calculations may also be performed using the QM/MM free energy perturbation (FEP) method. When applied to QM/MM the QM region is normally kept frozen and approximated by point charges representing the electrostatic potential of the QM electron density distribution during the MD runs to reduce computational cost, and so only the entropic contribution from the MM environment is measured. The entropic effect of the QM part is
estimated at energy minima and transition states using harmonic vibrational frequency calculations. The QM/MM-FEP method is now available as a module in ChemShell, which automates much of the lengthy setup process.

FEP calculations are normally based on top of mechanistic QM/MM investigations with the aim of obtaining a more realistic energy profile. As such, they have been combined with additional high level ab initio calculations, for example, local coupled cluster, for purely organic systems such as the hydroxylation reaction in p-hydroxybenzoate hydroxylase, as well as transition metal containing enzymes like aldehyde oxidoreductase.

**Excited-State Calculations**

Given the methodology for ground states described above, obtaining QM/MM energies of excited states using methods like TD-DFT and CASSCF/CASPT2 is straightforward, providing these methods are available in an interfaced QM code. With the availability of gradients for excited states, the exploration of an exited-state energy surface becomes possible. Therefore, vertical and adiabatic excitation energies as well as excited state minima can be obtained, as has been demonstrated by several groups. Publications working on flavin-based blue light photoreceptors including the Blue Light Using Flavin-Adenin-Dinucleotide and the Light-Oxygen-Voltage domain provide an insight into what has been done in this area.

Although studies of a single excited electronic state can be carried out using conventional methods, reactions that involve changes of electronic state require more sophisticated techniques. The conical intersection optimization methods in DL-FIND, for example, optimize the position where two electronic states intersect, which acts as a funnel for ultrafast relaxation to the lower state. ChemShell’s data structures, interfaces and QM/MM routines have been extended in order to support QM/MM multiple state calculations, including the calculation of interstate coupling gradients, and used to investigate the behavior of retinal models in solution. QM/MM conical intersection optimizations have subsequently been used to investigate the energy surfaces of DNA bases in solution and the DNA environment.

The multiple state interface is also being used to implement a surface hopping MD module for ChemShell based on the surface hopping routines in the MNDO semiempirical package. A development version of this module has been used to simulate the dynamical behavior of DNA bases in aqueous solution and the DNA environment.

**HIGH-PERFORMANCE COMPUTING**

There are two levels of parallelism possible in a ChemShell calculation: firstly, at the level of the energy or gradient evaluation, in which parallelism in external codes can be exploited; and secondly, at the higher level of tasks such as geometry optimization, which are handled by ChemShell.

ChemShell is parallelized using the message passing interface (MPI) formalism. It is not necessary to build ChemShell in parallel in order to exploit the parallelism of separate external programs, as they can simply be launched in parallel. However, running two separate binaries in an HPC environment can be inefficient and it is often not even possible to launch parallel executables in this way. ChemShell therefore offers the option to directly link some external codes as libraries so they behave as one parallel executable. This requires some modification to the external codes and so not all interfaces support this mode of operation. The second level of parallelism has been implemented in ChemShell using a task-farming approach. In task-farming parallelism, the overall set of MPI processors is split into smaller workgroups, each of which is assigned tasks such as energy evaluations. In a finite difference Hessian calculation, for example, each finite difference evaluation is independent and so may be calculated simultaneously using task-farming. Task-farming requires further modifications to external codes in order to share and make use of the MPI communicator handle that ChemShell uses to identify workgroups. The interfaces that support task-farming via a directly linked interface are listed in Table 1. It also possible in principle to perform task-farming calculations with other codes by launching separate binaries, but again, this is platform dependent.

Task-farming may also be used with the DL-FIND optimizer for algorithms where many independent energy evaluations are required, for example, the nudged elastic band method and population-based optimization using a genetic algorithm. The task-farm parallel nudged elastic band method has been used to investigate the mechanisms of hydrogen adsorption.
and carbon dioxide activation on zinc oxide surfaces, precursor steps in the industrial synthesis of methanol.

**CONCLUSION**

The basic methodology of QM/MM is now well established and has proven to be an efficient way of performing large-scale simulations in biology, chemistry, and materials science. In this article, we have detailed the flexible, modular environment for QM/MM calculations provided by ChemShell, allowing users to mix and match QM and MM packages through a consistent interface. Applications using the established methodology have demonstrated the wide variety of systems that QM/MM can be applied to, whereas recent developments in ChemShell highlight the continuing refinement of the QM/MM model to treat the environment more accurately, calculate free energies and study excited state reactions, and the use of QM/MM data to improve the crystallographic refinement process. Using the integrated DL-FIND optimizer, ChemShell has access to a range of powerful algorithms for exploring potential energy surfaces, and both codes have been modified to take advantage of massively parallel computing technology.

For further information about ChemShell and instructions for obtaining the code, visit the website www.chemshell.org.

**ACKNOWLEDGMENTS**

The ChemShell project was started by Paul Sherwood. ChemShell development is led by Thomas Keal and Paul Sherwood at STFC Daresbury Laboratory in collaboration with Alexey Sokol and Richard Catlow at University College London and the group of Walter Thiel at the Max Planck Institut für Kohlenforschung, Mülheim an der Ruhr, Germany. Other major code contributors include Alex de Vries, Alex Turner, Salomon Billiter, Stephan Thiel, Hans Martin Senn, Johannes Kästner, Tobias Benighaus, Eduardo Fabiano, Ya-Wen Hsiao, Yan Zhang, Eliot Boulanger, Ragnar Björnsson, John Buckeridge, and Andrew Logsdail. The authors of the DL-FIND optimization library are Johannes Kästner, Thomas Keal, Joanne Carr, and Judith Rommel.

We are grateful for financial support from EPSRC, the EU, Shell KSLA (Amsterdam), and the Volkswagenstiftung.

**REFERENCES**

1. Sherwood P, de Vries AH, Guest MF, Schreckenbach G, Catlow CRA, French SA, Sokol AA, Bromley ST, Thiel W, Turner AJ, et al. QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis. *Theochem J Mol Struct* 2003, 632:1–28.

2. Senn HM, Thiel W. QM/MM methods for biological systems. *Top Curr Chem* 2007, 268:173–290.

3. Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Angew Chem Int Ed* 2009, 48:1198–1229.

4. French SA, Sokol AA, Catlow CRA, Sherwood P. The growth of copper clusters over ZnO: the competition between planar and polyhedral clusters. *J Phys Chem C* 2008, 112:7420–7430.

5. Wu Q, Deng C, Peng Q, Niu Y, Shuai Z. Quantum chemical insights into the aggregation induced emission phenomena: a QM/MM study for pyrazine derivatives. *J Comput Chem* 2012, 33:1862–1869.

6. Smith W, Forester TR. DL_POLY_2.0: a general-purpose parallel molecular dynamics simulation package. *J Mol Graph* 1996, 14:136–141.

7. Shaik S, Cohen S, Wang Y, Chen H, Kumar D, Thiel W. P450 enzymes: their structure, reactivity, and selectivity—modeled by QM/MM calculations. *Chem Rev* 2010, 110:949–1017.

8. Li C, Shaik S. How do perfluorinated alkanoic acids elicit cytochrome P450 to catalyze methane hydroxylation? An MD and QM/MM study. *RSC Adv* 2013, 3:2995–3005.

9. Sen K, Hackett JC. Peroxo—iron mediated deformylation in sterol 14α-demethylase catalysis. *J Am Chem Soc* 2010, 132:10293–10305.

10. Usharani D, Zazza C, Lai W, Chourasia M, Waskell L, Shaik S. A single-site mutation (F429H) converts the enzyme CYP 2B4 into a heme oxygenase: a QM/MM study. *J Am Chem Soc* 2012, 134:4053–4056.
11. Senn HM, O’Hagan D, Thiel W. Insight into enzymatic C-F bond formation from QM and QM/MM calculations. *J Am Chem Soc* 2005, 127:13643–13655.

12. Metz S, Thiel W. A combined QM/MM study on the reductive half-reaction of xanthine oxidase: substrate orientation and mechanism. *J Am Chem Soc* 2009, 131:14885–14902.

13. Tuttle T. Averaging semiempirical NMR chemical shifts: dynamic effects on the subpicosecond time scale. *J Phys Chem B* 2009, 113:11723–11733.

14. Waller MP, Geethalakshmi KR, Bühl M. V-51 NMR chemical shifts from quantum-mechanical/molecular-mechanical models of vanadium bromoperoxidase. *J Phys Chem B* 2008, 112:5813–5823.

15. Geethalakshmi KR, Waller MP, Thiel W, Bühl M. V-51 NMR chemical shifts calculated from QM/MM models of peroxy forms of vanadium haloperoxidases. *J Phys Chem B* 2009, 113:4456–4465.

16. Frähmcke JS, Wanko M, Phatak P, Mroginski MA, Elstner M. The protonation state of Glu181 in rhodopsin revisited: interpretation of experimental data on the basis of QM/MM calculations. *J Phys Chem B* 2010, 114:11338–11352.

17. Schöneboom JC, Neese F, Thiel W. Toward identification of the Compound I reactive intermediate in cytochrome P450 chemistry: a QM/MM study of its EPR and Mössbauer parameters. *J Am Chem Soc* 2005, 127:5840–5853.

18. Sherwood P, de Vries AH, Collins SJ, Greatbanks SP, Burton NA, Vincent MA, Hillier IH, et al. Computer simulation of zeolite structure and reactivity using embedded cluster methods. *Faraday Disc* 1997, 106:79–92.

19. De Vries AH, Sherwood P, Collins SJ, Rigby AM, Rigutto M, Kramer GJ. Zeolite structure and reactivity by combined quantum–chemical–classical calculations. *J Phys Chem B* 1999, 103:6133–6141.

20. Björnsson R, Bühl M. Modeling molecular crystals by combined quantum-mechanical/molecular mechanical models of vanadium bromoperoxidase. *J Phys Chem B* 2008, 112:5813–5823.

21. Sokol AA, Bromley ST, French SA, Catlow CRA, Sherwood P. Hybrid QM/MM embedding approach for the treatment of localized surface states in ionic materials. *Int J Quant Comp* 2004, 99:695–712.

22. Gale JD, Rohl AL. The General Utility Lattice Program (GULP). *Mol Simul* 2003, 29:291–341.

23. Catlow CRA, French SA, Sokol AA, Al-Sunaidi AA, Woodley SM. Zinc oxide: a case study in contemporary computational solid state chemistry. *J Comput Chem* 2008, 29:2234–2249.

24. Goumans TPM, Catlow CRA, Brown WA, Kästner J, Sherwood P. An embedded cluster study of the formation of water on interstellar dust grains. *Phys Chem Chem Phys* 2009, 11:5431–5436.

25. Kästner J, Carr JM, Keal TW, Thiel W, Wander A, Sherwood P. DL-FIND: an open-source geometry optimizer for atomistic simulations. *J Phys Chem A* 2009, 113:11856–11865.

26. Metz S, Thiel W. QM/MM studies of xanthine oxidase: variations of cofactor, substrate, and active-site Glu802. *J Phys Chem B* 2010, 114:1506–1517.

27. Keal TW, Wanko M, Thiel W. Assessment of semiempirical methods for the photoisomerisation of a protonated Schiff base. *Theor Chem Acc* 2009, 123:145–156.

28. Lan Z, Fabiano E, Thiel W. Photoinduced nonadiabatic dynamics of 9H-guanine. *ChemPhysChem* 2009, 10:1225–1229.

29. Billeter SR, Turner AJ, Thiel W. Linear scaling geometry optimisation and transition state search in hybrid delocalised internal coordinates. *Phys Chem Chem Phys* 2000, 2:2177–2186.

30. Kästner J, Thiel S, Senn HM, Sherwood P, Thiel W. Exploiting QM/MM capabilities in geometry optimization: a microiterative approach using electrostatic embedding. *J Chem Theory Comput* 2007, 3:1064–1072.

31. Rommel JB, Goumans TPM, Kästner J. Locating instantons in many degrees of freedom. *J Chem Theory Comput* 2011, 7:690–698.

32. Gunn DSD, Allan NL, Foxhall H, Harding JH, Purton JA, Smith W, Stein MJ, Todorov IT, Travis KP, et al. Novel potentials for modelling defect formation and oxygen vacancy migration in Gd2Ti2O7 and Gd2Zr2O7 pyrochlores. *J Mater Chem* 2012, 22:4675–4680.

33. Nanda KD, Beran GJO. Prediction of organic molecular crystal geometries from MP2-level fragment quantum mechanical/molecular mechanical calculations. *J Phys Chem* 2012, 137:174106.

34. Benigna T, Thiel W. Efficiency and accuracy of the generalized solvent boundary potential for hybrid QM/MM simulations: implementation for semiempirical hamiltonians. *J Chem Theory Comput* 2008, 4:1600–1609.

35. Benigna T, Thiel W. A general boundary potential for hybrid QM/MM simulations of solvated biomolecular systems. *J Chem Theory Comput* 2009, 5:3114–3128.

36. Benigna T, Thiel W. Long-range electrostatic effects in QM/MM studies of enzymatic reactions: application of the solvated macromolecule boundary potential. *J Chem Theory Comput* 2011, 7:238–249.

37. Senn HM, Kästner J, Breiding J, Thiel W. Finite-temperature effects in enzymatic reactions—insights from QM/MM free-energy simulations. *Can J Chem* 2009, 87:1322–1337.
38. Senn HM, Thiel S, Thiel W. Enzymatic hydroxylation in \( p \)-hydroxybenzoate hydroxylase: a case study for QM/MM molecular dynamics. *J Chem Theory Comput* 2005, 1:494–505.

39. Boulanger E, Thiel W. Solvent boundary potentials for hybrid QM/MM computations using classical drude oscillators: a fully polarizable model. *J Chem Theory Comput* 2012, 8:4527–4538.

40. Hsiao Y-W, Sanchez-Garcia E, Doerr M, Thiel W. Quantum refinement of protein structures: implementation and application to the red fluorescent protein DsRed.M1. *J Phys Chem B* 2010, 114:15413–15423.

41. Kästner J, Senn HM, Thiel S, Otte N, Thiel W. QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: application to an enzymatic reaction. *J Chem Theory Comput* 2006, 2:452–461.

42. Mata RA, Werner H-J, Thiel S, Thiel W. Toward accurate barriers for enzymatic reactions: QM/MM case study on \( p \)-hydroxybenzoate hydroxylase. *J Chem Phys* 2008, 128:025104.

43. Metz S, Wang D, Thiel W. Reductive half-reaction of aldehyde oxidoreductase toward acetaldehyde: a combined QM/MM study. *J Am Chem Soc* 2009, 131:4628–4640.

44. Dieterich JM, Werner H-J, Mata RA, Metz S, Thiel W. Reductive half-reaction of aldehyde oxidoreductase toward acetaldehyde: ab initio and free energy quantum mechanical/molecular mechanical calculations. *J Chem Phys* 2010, 132:035101.

45. Sadeghian K, Bocola M, Schütz M. A conclusive mechanism of the photoinduced reaction cascade in blue light using flavin photoreceptors. *J Am Chem Soc* 2008, 130:12501–12513.

46. Sadeghian K, Bocola M, Schütz M. A QM/MM study on the fast photocycle of blue light using flavin photoreceptors in their light-adapted/active form. *Phys Chem Chem Phys* 2010, 12:8840–8846.

47. Salzmann S, Silva-Junior MR, Thiel W, Marian CM. Influence of the LOV domain on low-lying excited states of flavin: a combined quantum-mechanics/molecular-mechanics investigation. *J Phys Chem B* 2009, 113:15610–15618.

48. Lan Z, Lu Y, Fabiano E, Thiel W. QM/MM nonadiabatic decay dynamics of 9H-adenine in aqueous solution. *ChemPhysChem* 2011, 12:1989–1998.

49. Lu Y, Lan Z, Thiel W. Monomeric adenine decay dynamics influenced by the DNA environment. *J Comput Chem* 2012, 33:1225–1235.

50. Fabiano E, Keal TW, Thiel W. Implementation of surface hopping molecular dynamics using semiempirical methods. *Chem Phys* 2008, 349:334–347.

51. Lu Y, Lan Z, Thiel W. Hydrogen bonding regulates the monomeric nonradiative decay of adenine in DNA strands. *Angew Chem Int Ed* 2011, 50:6864–6867.

52. Keal TW, Sherwood P, Dutta G, Sokol AA, Catlow CRA. Characterization of hydrogen dissociation over aluminium-doped zinc oxide using an efficient massively parallel framework for QM/MM calculations. *Proc R Soc A* 2011, 467:1900–1924.

53. Dutta G, Sokol AA, Catlow CRA, Keal TW, Sherwood P. Activation of Carbon Dioxide over Zinc Oxide by Localised Electrons. *ChemPhysChem* 2012, 13:3453–3456.