QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis. J Mol Struct: THEOCHEM 632(1–3):1–28

ARTICLE *in* JOURNAL OF MOLECULAR STRUCTURE THEOCHEM · AUGUST 2003

Impact Factor: 1.37 · DOI: 10.1016/S0166-1280(03)00285-9

ELSEVIER

Journal of Molecular Structure (Theochem) 632 (2003) 1–28

THEO CHEM

www.elsevier.com/locate/theochem

QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis

Paul Sherwood[a,*], Alex H. de Vries[a], Martyn F. Guest[a], Georg Schreckenbach[a], C. Richard A. Catlow[b], Samuel A. French[b], Alexey A. Sokol[b], Stefan T. Bromley[b,1], Walter Thiel[c], Alex J. Turner[c], Salomon Billeter[c], Frank Terstegen[c], Stephan Thiel[c], John Kendrick[d], Stephen C. Rogers[d], John Casci[e], Mike Watson[e], Frank King[e], Elly Karlsen[f], Merethe Sjøvoll[f], Adil Fahmi[f], Ansgar Schäfer[g], Christian Lennartz[g]

[a]*CLRC Daresbury Laboratory, Warrington WA4 4AD, UK*
[b]*Davy Faraday Research Laboratory, The Royal Institution of Great Britain, 21 Albemarle Street, London W1S 4BS, UK*
[c]*Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim, Germany*
[d]*Imperial Chemical Industries PLC, Wilton, Teeside TS90 8JE, UK*
[e]*Synetix, Belasis Avenue, Billingham TS23 1LB, UK*
[f]*Norsk Hydro Oil and Energy Research Centre, Section for Hydrocarbon Processes and Catalysis, 3907 Porsgrunn, Norway*
[g]*BASF Aktiengesellschaft ZX/ZC, C13 67056 Ludwigshafen, Germany*

Received 19 November 2002; accepted 2 December 2002

Abstract

We describe the work of the European project QUASI (Quantum Simulation in Industry, project EP25047) which has sought to develop a flexible QM/MM scheme and to apply it to a range of industrial problems. A number of QM/MM approaches were implemented within the computational chemistry scripting system, ChemShell, which provides the framework for deploying a variety of independent program packages.

This software was applied in several large-scale QM/MM studies which addressed the catalytic decomposition of N_2O by Cu-containing zeolites, the methanol synthesis reaction catalysed by Cu clusters supported on ZnO surfaces, and the modelling of enzyme structure and reactivity.

Keywords: Catalysis; ChemShell; QUASI; Embedding

1. Introduction

The QUASI project (Quantum Simulation in Industry) was funded by the EU, project EP25047 under the High Performance Computing and Networking (HPCN) action of Framework 4.

* Corresponding author.
E-mail address: P.Sherwood@dl.ac.uk (P. Sherwood).
[1] Present address: Laboratory of Applied Organic Chemistry and Catalysis, Delftchem Tech, Julianalaan 136, 2628 BL Delft, The Netherlands.

0166-1280/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0166-1280(03)00285-9

The project aimed to build on work done within the EUROPORT [1] project of Framework 3, in which a wide range of scientific and technical packages were ported to parallel machines. The QUASI project moved the focus from individual code parallelisation to the parallel implementation of integrated modelling approaches, in particular the QM/MM method which couples quantum and classical simulation techniques [2–6].

The project included two partners from EUROPORT: CLRC Daresbury Laboratory, the developers of the ab initio program package GAMESS-UK [7], and Walter Thiel's group, then at the University of Zürich, developers of the semiempirical package, MNDO [8]. These teams, with a focus on quantum molecular simulation, were joined by Richard Catlow's group at the Royal Institution which provided expertise in atomistic simulation and embedded cluster approaches to solid-state and surface problems. Modelling teams from three industrial partners, Norsk Hydro, ICI (Synetix) and BASF, brought real life problems in the areas of zeolite, surface and enzyme catalysis.

The main objectives of the QUASI project were to

- address some of the barriers to the uptake of QM/MM modelling approaches in mainstream modelling, including validation work on the QM/MM methods already in use in the partner groups, and development of geometry optimisation approaches suitable for complex systems;
- apply the software developed to a range of real industrial problems;
- demonstrate the potential for high performance computing within the industrial molecular modelling arena.

2. Methodology and implementation

2.1. ChemShell—a modular environment for computational chemistry

ChemShell [9] was developed in response to a variety of challenges encountered when trying to use UNIX shell programming techniques to integrate a range of computational chemistry packages to perform embedded cluster and QM/MM calculations. Although suitable for simple tasks, the UNIX shell script approach proved increasingly awkward as more complex tasks were tackled. Many tasks required the use of a number of packages, together with utility programs to convert formats and create inputs, leading to a number of problems. In particular,

- there was a need to introduce some control logic, loops, etc. at the higher level, and standard shell programming languages, such as csh and ksh were not well suited for coding complex algorithms;
- there were performance implications, particularly on parallel computers, if many individual programs had to read and write many files, which (for macromolecular systems) could be quite large in size;
- the control data would generally be distributed amongst many files (each pertaining to a different program) and it was therefore quite easy to lose track of the important data. This problem was exacerbated if the same program was used many times on different systems.

To address these issues it was decided to choose one of the freely available interpreted languages, Tool Command Language (Tcl) [10] and integrate the computational chemistry programs to enable the packages and utilities to be accessed from a uniform procedural interface. A Tcl script would then control the sequence of steps, some of which would be executed as Tcl procedures, some by code directly linked into the interpreter and some by system calls to execute independent packages. The packages could be integrated in such a manner that the user input was independent, as far as possible, from the particular third-party code that was used to perform the computation, allowing more general scripting of algorithms.

2.1.1. The ChemShell object types

Since Tcl works entirely with strings, and conversion of chemical data to and from string representation is potentially inefficient, ChemShell implements, in C code, a number of classes of data objects that are then referenced by the scripts using short naming strings or tags. These objects are stored as files between jobs but in favourable cases can be held in memory (cached) during job execution to minimise IO activity. The object types implemented in ChemShell are given in Table 1.

Table 1
The ChemShell data object types

Object type	Attributes
Fragment	Molecular coordinates, crystallographic symmetry. Also includes shells and point charges when used
Z-Matrix	Internal coordinate representation of a molecule, including variables and constants (and optionally some atoms specified by Cartesian coordinates)
Field	Numerical data on a regular or irregular spacial grid
Matrix	Numerical data array
Graph	OpenGL primitives for visualisation

2.1.2. Syntactical features

Tcl has very little syntactical structure and each statement is interpreted simply as a list of words; the first word in each case is assumed to be a command and the remaining words are the associated arguments. Braces { }, are used to group words into lists, and [ ] to substitute a list of words with the result of treating the list as a command. ChemShell scripts follow this approach, but two specific features were added:

1. The equals sign (=) is used to provide arguments as name/value pairs, e.g.

```
energy theory=mndo coords=benzene.coords energy=benzene.energy
```

   where `benzene.coords` and `benzene.energy` are names of ChemShell objects, the first (a Fragment object) is the source of the structure, and the second (a Matrix object, in this case containing a single element) holds the resulting energy.
2. The colon (:) is used after a name/value pair to signify that the following argument, or list of arguments, is to be processed by the ChemShell module specified by the preceding argument e.g.

```
energy theory=gamess : { basis=dzp hamiltonian=b3lyp} coords=benzene.c \
       energy=benzene.energy
```

   Here the basis and Hamiltonian arguments are to be interpreted by the gamess module (and not by the energy module), enabling it to be completely independent of the source of the energy to be used.

2.1.3. Utility modules

ChemShell incorporates code for a variety of computational chemistry operations including:

- *z*-matrix manipulation (including modifying variables and constants to map out potential energy surfaces);
- merging/splitting of chemical structures;
- least-squares fitting of two structures;
- least-squares fitting of charges to electrostatic potentials;
- vibrational frequencies by finite difference, often limited in QM/MM studies to a small active region.

2.1.4. Graphical interfaces

ChemShell has an internal graphics module, controlled by Tcl scripts or by a simple Tcl/Tk GUI and this functionality has been used within QUASI for post-calculation analysis, e.g. the visualisation of electrostatic potentials and vibrational modes. The design of ChemShell is based on the idea that maximum flexibility can be achieved by setting up the calculations using Tcl scripts. However, especially in the industrial modelling context targeted by QUASI, graphical interfaces are now widely adopted and we have therefore developed, within QUASI, a fully functioning GUI for use in the Cerius2 environment from Accelrys Inc. We are currently working to

make this same functionality available within a number of open source modelling environments as well, e.g. the PyMOL package [11], which shows considerable promise for biomolecular applications.

2.2. Some ChemShell modules

In this section we review some of the more important modules that are available in ChemShell and which are used in the QM/MM studies performed by the QUASI project.

2.2.1. Interfaces to classical and quantum simulation codes

DL_POLY. The classical modelling of covalent materials has been performed with a number of Fortran subroutines taken from the DL_POLY [12] package by Bill Smith. DL_POLY is a general purpose molecular dynamics package and it has been applied to a wide variety of chemical systems [13]. However, it does not, at present have any internally stored force fields and in order to use the package in the framework of a scripted system we have added code to prepare some of the DL_POLY data files. The most important such development is a set of routines to generate the force-field or topology file (FIELD) from a concise listing of molecular mechanics parameters, the molecular atom types and connectivity. This algorithm also includes code to assign the force field atom types based on the connectivity and character of the connected atoms. We have extended the range of functional forms supported to incorporate the cross-terms needed for the CFF force field derived by Hill and Sauer for zeolites [14,15].

It was not considered practical to code the rules required to set up biomolecular force fields, such as CHARMM and AMBER, partly because of the complexity of these force fields and, in particular, the fact that the assignment of parameters is based on libraries of residues, rather than simply processing atom types and connectivity. Rather than build these residue topology databases and assign the parameters ourselves, we have chosen to load the parameter and topology (PSF) files from CHARMM [16] and convert them for use with DL_POLY.

GULP. GULP (the General Utility Lattice Program) from Julian Gale [17] is used as the classical component of our simulations when shell-model [18] force fields are employed. The code is linked directly into the ChemShell binary, and includes a small number of modifications with respect to standard GULP 1.3. Most important is the fact that when the shell relaxation step is performed a term arising from the interaction between the shell charge and the electrons and nuclei of the quantum mechanical system may be incorporated.

CHARMM. The CHARMM macromolecular simulation package [16] serves two roles in the ChemShell environment. It is interfaced as a package in its own right, enabling the energy and forces computed within CHARMM to be deployed in the geometry optimisation and dynamics algorithms provided in ChemShell. However, since the coupling between ChemShell and CHARMM is via UNIX pipes the data transfer rates are not high enough for the ChemShell/CHARMM system to compete with a simulation using CHARMM's own dynamics and optimisation drivers. In the QUASI project we have used CHARMM software to define the force field which can then subsequently be used by the DL_POLY molecular mechanics routines.

GROMOS. An interface to GROMOS96 has been developed with support for QM/MM calculations by code to implement the adjustments to the classical force field terms at the QM/MM boundary. In this context, sections of the GROMOS96 package [19] have been combined into a separate program that performs single-point calculations returning the MM energy and gradient contributions for a given geometry. A Tcl interface provides the input to this program essentially by passing the Cartesian coordinates and the required atom attributes. Data transfer occurs via UNIX pipes and is synchronized by UNIX sockets. This implementation cannot compete in speed with the standalone GROMOS96 package for purely classical MM minimisations and molecular dynamics runs, but it allows an easy access to the GROMOS force field in QM/MM work with ChemShell. AM1/GROMOS96 geometry

optimisations for *p*-hydroxybenzoate hydroxylase have been performed [20] with this implementation.

GAMESS-UK. GAMESS-UK (Generalised Atomic and Molecule Electronic Structure System [7]) is an ab initio quantum mechanics program supporting SCF, DFT and wide range of correlated methods and properties. Over recent years a number of embedding methods have been implemented with GAMESS-UK, e.g. [21–23] and various enhancements have been made to support a range of electrostatic models. The code has a range of SCF convergence methods which have been further refined within QUASI to deal with some of the more problematic application areas such as the copper-containing clusters discussed below. The SCF, DFT and MP2 sections of the code have been parallelised and show good scalability on a range of parallel platforms. To support execution on massively parallel machines (Section 2.4) GAMESS-UK can be compiled as a library and linked into the ChemShell executable.

MNDO. The MNDO code [8,24] is a semiempirical QM program that offers a wide range of functionality. It provides all the standard semiempirical methods (e.g. MNDO, AM1, PM3, and MNDO/d) as well as the recently developed methods with orthogonalisation corrections (OM1, OM2). Electron correlation can be treated explicitly by perturbation theory or various forms of configuration interaction (CI) up to full CI within a given active space (GUGACI formalism). Analytic gradients are available for most methods both at the SCF and the CI level, while an analytic Hessian can be computed only for MNDO-type methods at the SCF level. Concerning QM/MM approaches, the code allows for calculations with external point charges, and it can determine both the electrostatic potential and the electric field at these external points (for several semiempirical conventions) such that different QM/MM coupling schemes [25] are supported (A–D, see Section 2.3.4). In addition, both link atoms and connection atoms [26] can be chosen for terminating bonds that have been cut at the QM/MM boundary.

Within ChemShell, MNDO is run as a stand-alone binary, exploiting shared-memory parallelism when running on a multi-processor system.

Turbomole. TURBOMOLE [27] is a highly optimised ab initio package designed for workstation-class computers. Within ChemShell, the TURBOMOLE interface supports HF and DFT wavefunctions. Serial as well as MPI-parallel versions are supported. The interface passes commands to the TURBOMOLE define module to create the necessary input files for a TURBOMOLE run and then starts the corresponding modules (dscf/grad or ridft/rdgrad) which run as stand-alone binaries. QM/MM geometry optimisations and molecular dynamics runs benefit from the ability of the TURBOMOLE gradient modules to compute analytical forces quickly also for large arrays of external point charges. Of particular value to QUASI has been the efficient RI-DFT implementation which has been used, for example, in the QM/MM studies of TIM [28] described below.

Other quantum mechanical interfaces. Interfaces are also available for other packages, including MOPAC, CADPAC and Gaussian98. However, these do not support the full range of QM/MM methods presented in this paper and will not be discussed further here.

2.2.2. Geometry optimisation

ChemShell incorporates a number of geometry optimisation algorithms. Two will be described here. The first, NewOpt, is designed for small to medium sized molecules, while HDLCOpt was developed within the QUASI project for macromolecular systems.

NewOpt. Newopt is a (Tcl) script-driven geometry optimiser for reactions of small molecules (gas phase, and absorbed on substrates). It works in mixed Cartesian/*z*-matrix coordinates and provides the functionality expected in quantum chemistry codes, including implementations of a wide variety of algorithms, such as BFGS, P-RFO and DIIS. It incorporates a number of specialised features developed for studying embedded reactive sites, for example, finite difference Hessians can be computed in the eigenmode coordinate system. In this way a partial finite difference Hessian calculation can be used to refine the information pertaining to particular modes without the cost of a full Hessian calculation. Adaptations of the P-

RFO method to deal with modes with near-zero eigenvalues, corresponding to low energy translations and rotations of the adsorbate, have been developed.

HDLCOpt. In macromolecular systems with thousands of atoms, energy minimizations are often carried out by first-order methods that only require the energy and gradient (e.g. steepest descent or conjugate gradient methods). Second-order methods which include Hessian information are inherently more efficient for energy minimization and particularly for transition state searches. However, the computational effort for standard second-order optimisation algorithms rises quickly with molecular size and can become quite significant for macromolecular systems, both in terms of CPU and memory requirements. Therefore, it is desirable to implement linear scaling approaches also for geometry optimisers. One strategy is to exploit sparsity in the usual second-order optimisation algorithms [29–32]. We have followed a different strategy and have implemented a divide-and-conquer approach [33] in ChemShell.

Briefly, the system is partitioned into user-defined fragments (e.g. amino acid residues in a protein), and all coordinate manipulations required by second-order algorithms are performed exclusively within these fragments (which ensures linear scaling). From a Cartesian input geometry, the code automatically generates a redundant set of primitive internal coordinates in each fragment and constructs the corresponding non-redundant set of delocalised internal coordinates [34]. The relative orientation of the fragments is best described by Cartesian coordinates, and therefore the redundant set of primitive coordinates is supplemented with the corresponding Cartesian coordinates for each fragment and then delocalised, which yields a non-redundant set of hybrid delocalised internal coordinates (HDLC) [33]. Other coordinate options are also available. In particular, there are automatic schemes for recovery from coordinate breakdown during HDLC construction and procedures for imposing constraints for selected coordinates [33].

The actual optimisation utilizes a limited-memory quasi-Newton algorithm (L-BFGS) for energy minimization and a Hessian eigenmode-following algorithm (P-RFO) for the transition state search. Microiterative schemes [35–38] are implemented to allow for alternating optimisation of a core region and the environment, which may, but need not, coincide with the QM and MM regions, respectively. Such a microiterative divide-and-conquer approach is particularly advantageous for locating transition states because it enables us to search uphill (P-RFO) in a suitably chosen fragment for the reactive core and to minimize downhill (L-BFGS) in all other fragments. In this case, the Hessian for the P-RFO algorithm needs to be computed only for the relatively small reactive core.

Due to the modular structure of ChemShell the HDLCOpt code can be used not only with QM/MM methods, but also with pure QM and pure MM methods. The initial tests [33] have demonstrated its efficiency for energy minimizations of thrombin with crystal water (5445 atoms) using the CHARMM 23f6 all-atom force field, and for transition state searches in Diels-Alder and hydroxylation reactions at the semiempirical AM1 level. Recent applications of HDLCOpt within ChemShell include DFT studies of large organometallic systems [39] and QM/MM studies of enzymes (Section 3.3).

2.2.3. Molecular dynamics

The ChemShell MD module is based on routines from the molecular simulation code DL_POLY [12]. The functionality includes NVE, NVT and NPT integrators, handling of constraints by Shake and rigid body motion by quaternions. In keeping with the modular design, the MD module can perform QM, MM and QM/MM simulations using the full range of interfaced packages. The interface is object oriented, a dynamics object is created and subsequently manipulated by a series of method calls which manipulate the underlying data structures. The `configure` method allows the simulation conditions to be changed, an example of a simple simulation script is given below.

```
dynamics dyn1 coords=c theory=mndo temperature=300 timestep=0.005
# initialise velocities
dyn1 initvel
set nstep 0
while {$nstep < 10000 } {
  dyn1 force
  dyn1 step
  .....
  # additional Tcl commands here (e.g. restraint code)
  incr nstep
}
dyn1 configure ensemble=NVT temperature=300
# etc.
dyn1 destroy
```

The script approach allows greater flexibility in defining simulation protocols than a conventional driver, which would pass control to the MD loop implemented within, for example, a Fortran code. Tcl commands to access (and modify) all the state parameters of the simulation can be inserted in the loop. A simple example of an application of this is the addition of a user-defined restraint term (which can be implemented as a Tcl procedure).

2.3. Hybrid QM/MM coupling schemes

While the basic idea of coupling quantum and classical calculations is an obvious and well-established one, the implementation is fraught with choices and practical difficulties. Since a range of QM/MM approaches are included in the QUASI project, we start this section with a general discussion of the possible approaches and the way we have chosen to tackle the issues that arise.

2.3.1. Classification of QM/MM methods

The basic idea is to identify a small region, the Inner Region in Fig. 1(a) to which the higher (QM) level of theory is to be applied. In a covalent system the excision of this region will involve breaking bonds; in a more ionic material there will need to be some treatment for the nearby ions that are to be treated classically. Invariably it will be necessary to compute the QM wavefunction of the Inner Region in the presence of at least a layer of atoms which serve to provide a satisfactory termination for the QM calculation. We can classify QM/MM models according to the way this boundary region is handled:

- In many methods, traditionally those used for covalent systems, a number of additional atoms, which have no counterparts in the real system, are added to the QM region to terminate the broken bonds, these are typically called link atoms. We will refer to these methods as *link-atom methods.*

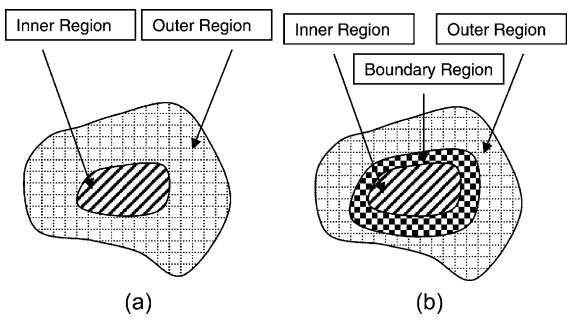

Fig. 1. Inner, outer and boundary regions.

- When handling ionic systems, the models, often referred to as embedding models, handle the termination of the QM region by defining a boundary region (Fig. 1(b)) in which the atoms have some quantum character, as well as being described fully by the classical force field. No new centres are needed. We will refer to these models as *boundary region methods*.

Within QUASI we have used link-atom methods for the modelling of zeolite and enzyme systems, and boundary region methods for modelling metal oxide surfaces.

2.3.2. Nature of the energy expression

QM/MM schemes can also be classified by the nature of the total energy expression. At one extreme, a QM/MM energy expression can be obtained by adding the QM and MM contributions, together with the appropriate coupling terms

$$E = E(\text{Inner}, \text{QM}) + E(\text{Outer}, \text{MM}) + E(\text{Couple}, \text{QM/MM}) \quad (1)$$

E(Couple, QM/MM) is a QM/MM coupling term that includes all interactions between the two regions, for example, classically handled bonding and van der Waals interactions and modifications to the QM Hamiltonian to reflect the electrostatic influence of some or all of the atoms in the outer region (vide infra).

We will refer to this approach as the *additive* QM/MM model. The method can be applied easily to systems where there are no boundary atoms (as in studies of solvation) but as presented above the approach takes no account of the handling of link atoms at the interface. An alternative approach, which we will term *subtractive*, applies the MM scheme to the whole (Inner + Outer) system and then corrects for the multiple counting by subtracting the energy (computed at the classical level) of the Inner system.

$$E = E(\text{Inner} + \text{Link}, \text{QM}) + E(\text{Inner} + \text{Outer}, \text{MM}) - E(\text{Inner} + \text{Link}, \text{MM}) \quad (2)$$

The subtraction removes the multiple counting of the energy of the Inner region, and also removes any contribution to the total energy from the link atoms. However, to do this it relies on the ability of the MM energy expression to closely approximate the QM potential energy surface for the Inner and Link regions.

In fact, these two energy expressions represent extremes and a number of intermediate models are possible. If both of the MM contributions to Eq. (2) are computed such that they exclude contributions arising solely from the atoms in the interior of the Inner region, the need for the force field to describe the whole Inner region is reduced, while still exploiting the subtractive approach to delete the terms associated with the link atoms. In the limit of excluding all Inner atoms one arrives at an additive expression with a link atom correction [25]

$$E = E(\text{Inner} + \text{Link}, \text{QM}) + E(\text{Outer}, \text{MM}) + E(\text{Couple}, \text{QM/MM}) - E(\text{Link}, \text{MM}) \quad (3)$$

Within QUASI we have used additive methods throughout, although the subtractive methods have also been implemented. The motivation for this choice is twofold:

1. The subtractive approaches include all coupling between the inner and outer regions at the MM level of theory and the QM calculation is performed effectively as a gas-phase cluster calculation. This rules out the computation of quantum mechanical properties (e.g. spectroscopic properties) of the inner region that are influenced by the environment;
2. The requirement that the force field accurately reproduces the QM potential energy surface implies extensive development work on force fields, especially for systems undergoing reactions. This is certainly tractable (see, for example, the work of Sauer and co-workers [38,40–42]) but it adds to the time required to start work on a new system, a major drawback in an industrial context.

We do not apply an explicit link atom correction as in Eq. (3). Instead we find that a simple and robust alternative is to retain the link atom energy but to remove from the MM force field energy expression those terms which would, if included, lead to multiple counting of contributions to the energy of particular distortions. We will return to this in Section 2.3.8.

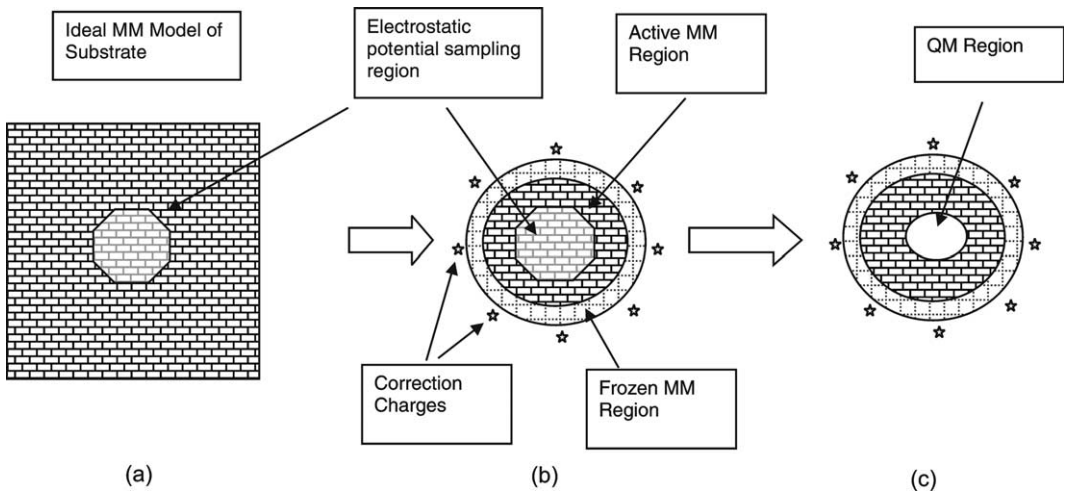

Fig. 2. Generation of finite cluster models.

2.3.3. Generation of finite cluster models

Many of the systems to which QM/MM models are applied are condensed phase in nature, and the simulation of such materials is often made tractable by the use of periodic boundary conditions. However, periodic boundary conditions are problematic for QM/MM simulations as many QM codes (e.g. GAMESS-UK and MNDO) do not contain the ability to compute the interaction of the QM system with an infinite charge distribution. We have implemented a charge fitting approach which enables a finite cluster to model accurately a single QM defect in an infinite, perfect classical crystalline system. In many cases, this is actually more pertinent to the real system under study than a model system with a periodically repeating defect.

The approach adopted is to start from the perfect, periodic system modelled using the classical force field approach, Fig. 2(a), and from it cut a cluster into which the QM model is to be embedded Fig. 2(b). A sampling grid is defined such that the points occupy the region in the interior of the classical cluster, including the region where the QM embedding is to take place and some of the surrounding MM region that will be subject to relaxation. The electrostatic potential on these points is computed (using the classical charge model) for both the perfect, infinite system and for the finite cluster (Fig. 2(a) and (b)). The difference in these potential datasets represents the error introduced by truncating the cluster. It is often possible to reduce the magnitude of this error by adjusting the charges on the atoms on the surface of the finite cluster. This is particularly true of covalent solids where an adjustment can be applied for each bond cleaved by the cluster generation.

The remaining error in the electrostatic potential is corrected by defining a set of additional point charges, placed around the outside of the cluster. The magnitudes of these charges are found by least-squares fitting to the error potential so that they compensate for the missing long-range electrostatic terms. This approach has been applied to both zeolite and metal oxide systems described below. Finally the QM region can be defined (and with it a boundary layer if required, see Fig. 2(c)).

2.3.4. Coupling of the QM and MM regions

The QM and MM regions are coupled by bonded and non-bonded interactions. The bonded terms are generally handled by terms in the classical force field or (for boundary region methods) by a combination of quantum and classical terms between QM and boundary atoms. The non-bonded terms comprise van der Waals and electrostatic interactions. The former are handled in the same way as they are in the MM region (perhaps with modified parameters). Bakowies and Thiel [25] defined three ways of treating the QM/MM electrostatic interaction, labelled A–C, as follows.

A. *Mechanical embedding*. The QM calculation is essentially performed in the gas phase, without electronic coupling to the environment. The electrostatic interaction between QM and MM regions is either omitted or performed by the MM code, using a classical point charge model for the QM charge distribution (e.g. a potential derived charge model).

B. *Electrostatic embedding*. The classical partition appears as an external charge distribution (e.g. a set of point charges) in the QM Hamiltonian. The polarisation of the QM region by the MM charge distribution thus occurs as part of the QM electronic structure calculation. The partial charges used to describe the MM distribution are frequently taken to be those used in the force field [43–45], relying on the use of electrostatic properties in the force field charge derivation (the work on enzymes in Section 3.3 follows this approach). When using an aluminosilicate CFF force field [14] for electrostatic QM/MM modelling of zeolites (Section 3.2), it was found necessary to replace the original MM charges with those derived by fitting to periodic quantum mechanical electrostatic potentials. In ab initio schemes it is clear that the electrostatic embedding scheme should be implemented, at least at long range, by adding the contribution of the MM point charges to the 1-electron Hamiltonian. However, within the semiempirical formalism the definition of the electrostatic potential is more ambiguous as a result of the overlap approximations used, and alternative formulations for the 1-electron integral terms have been suggested [44,46,47].

C. *Polarised embedding*. The polarisation of the MM region in response to the QM charge distribution is also included. Intuitively this makes most sense when the force field incorporates polarisation as unpolarised force-fields implicitly incorporate MM polarisation in their parameterisation, and care must be taken to ensure such implicit contributions do not occur in the QM/MM potential. A variety of models for classical polarisation are possible, including the shell model [18], and coupled distributed atomic polarizabilities [48]. Polarisation of MM atoms close to the QM region (e.g. those connected by link-atom terminated bonds) was found to be unphysically large, leading to the suggestion that these atoms be treated as unpolarizable [25].

In model C the result of the change in the MM charge distribution from the classical polarisation is not propagated to the QM calculation resulting in a non-variational total energy. For this reason the QUASI implementation uses model D below.

While not part of the original definition, model D [49] is defined as an extension to model C where QM and MM polarisations are made *self-consistent*, either by iterative solution of the SCF and polarizability problems [50–52], or by matrix inversion techniques, as exemplified by the Direct Reaction Field (DRF) model [53]. The iterative approach has been taken within QUASI and the implementation of a self-consistent QM/MM scheme based on shell model [18] potentials is described in Section 2.3.9.

2.3.5. Dividing the system into quantum and classical regions

The approach taken to the definition of the QM region depends on the nature of the force field model to be employed for the classical region. Various factors, such as the size of the QM region needed to provide an accurate model and the nature of the bonds cleaved, influence this choice. However, care must also be taken to ensure that removal of the group of atoms identified for treatment at the QM level leads to a sensible total charge on the remaining MM system. We will discuss here a number of approaches we have used to ensure this.

Charge-group approaches. Many biochemical force fields have the feature that sets of neighbouring atoms are grouped together such that the total charge on the group is an integer (usually zero) [54]. This is a convenient feature for a force field as it enables a molecule built by combining these charge groups to have an integral charge without any adjustment of the atomic charges being necessary. If the QM region is defined to be a suitably terminated neutral group (or a number of connected neutral groups) a neutral MM 'embedding hole' can be easily created (Fig. 3).

This approach has been adopted for our studies of enzymes.

Structural elements for ionic solids. In order to study a localised state in an extended insulator or semiconductor (a dielectric) we define two main objects: a reference (defect-less) system and a defect. Within the region of localisation of the defect state

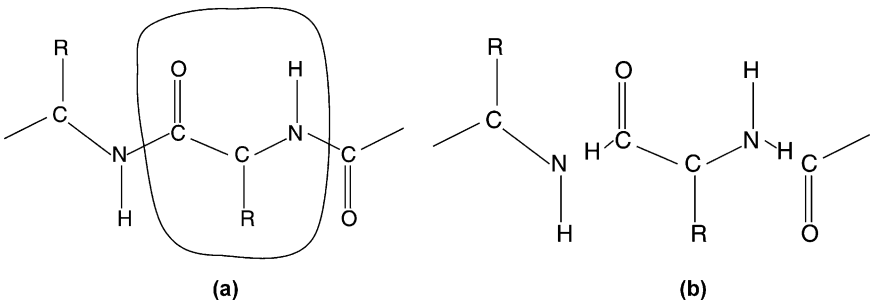

Fig. 3. Choosing the QM region based on a force field neutral group. (a) The pure MM system, showing the neutral group boundary, (b) the QM region with link atoms.

the local atomic and electronic structure (including, for instance, the charge and spin density) are significantly different from that of the reference system whereas outside, the dielectric closely resembles the reference system.

The reference system is postulated to consist of structural elements that weakly overlap, interact with each other and are polarised by external fields as proposed by Kantorovich [55–57]. Such structural elements are destroyed or strongly perturbed in the defect region (upon formation of structural defects, electron transfer, etc.) while staying intact in the system remainder. Each structural element is associated with an electron group, and the wave function of both reference and defective systems are represented by an antisymmetrised product of the corresponding group functions as introduced in the many electron theory of separability and localisation [58]. The choice regarding the division of a many electron system into electron groups is restricted by the provision that each structural element should comprise an integral number of electrons and electron correlation effects between electron groups are small. This approach underlies the embedding approach for modelling ionic solids detailed in Section 2.3.9, where spherical polarisable ions are treated as structural elements.

Semi-ionic solids. Within our calculations on zeolites (Section 3.2) a different scheme has been adopted. The aluminosilicate force field used [14] does not consist of charge groups, and although a neutral QM fragment (comprised of an integral number of SiO_2 units) could be defined, this would not be a natural choice on chemical grounds. A more obvious choice would be a number of silicon (or aluminium) atoms together with all of their attached oxygen atoms. However such a group will not be expected to have a zero charge, and an adjustment of the remaining MM charges is needed. The first stage of the approach is to adjust the charge on centre M_1 (see Fig. 5) to account for the deletion of the M_1-Q_1 bond. For an aluminosilicate force field with silicon charges of $+2x$ and oxygen charges of $-x$ (Fig. 4), each Si–O bond can be considered to contribute a charge of magnitude $0.5x$ to the atom at each end (since there are two bonds to each O and 4 to each Si). The idea of deriving electrostatic models from a sum of dipolar contributions from the bonds to each atom has been used for more general chemical systems, for example, the MM3 force field [59] uses an electrostatic model including bond dipole terms. For each QM/MM bond that each atom M_1 is involved in, the charge on M_1 is reduced by $0.5x$, thus ensuring that the defect created in the MM lattice is electrically neutral.

2.3.6. Charge adjustments at the QM/MM boundary

With all of the polarised QM/MM schemes (models B–D in Section 2.3.4) problems can be expected when point charges modelling the MM region closely approach the QM region. In the absence of link atoms, close approach is usually

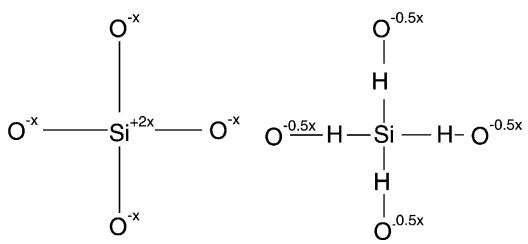

Fig. 4. Adjusting the classical charges around the QM region.

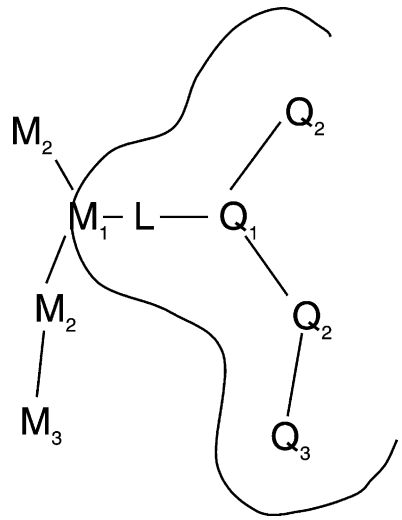

Fig. 5. Atom labelling in the QM/MM link region.

prevented by the non-bonded interaction potential which is repulsive at short range. However in the region of bonds across the QM/MM boundary, some adjustment to the classical charge distribution is essential, as the nearest point charges to the QM region will be at most a single bond distance away. In the case of terminating link atoms (e.g. hydrogen) the link atom will be almost superimposed on the first classical atom (M_1 in Fig. 5).

Clearly this problem will be more severe when large basis sets are used, and in fact it is possible to disregard it in the semiempirical case [44]. Antes and Thiel [49,60] have discussed a variety of approaches to the problem and suggested the L1–L3 classification used below.

Selective deletion of one-electron integral terms. For QM calculations with small basis sets the leading spurious interaction is that of the basis functions on the link atom with the classical MM charges. Since the link atom is an artefact of the QM/MM scheme it has been suggested that the model can be improved simply by deleting the 1-electron Hamiltonian contributions (Eq. (4)) which involve basis functions i or j sited on the link atom l (scheme L1 [49])

$$V_{ij}^{A} = \langle \phi_i(l) | \frac{-q_A}{r_{lA}} | \phi_j(l) \rangle \tag{4}$$

The L1 model is found to be effective for semiempirical wavefunctions but the difference in potential acting on nearby basis functions causes unphysical polarisation for ab initio QM models. This model is available in the ChemShell QM/MM implementation when MNDO is chosen as the QM code. Antes and Thiel also defined a further scheme, denoted L3, in which all 1-electron terms are included with the exception of those for which functions i and j are located on the link atom and q_A is the neighbouring MM atom (M_1), but as this scheme did not offer consistent performance, it has not been implemented in the QUASI software.

Deletion of selected atomic charges. Perhaps the simplest way of dealing with the charge on the nearby classical centres is to delete them from the Hamiltonian. The QUEST Gaussian/AMBER scheme [43] omitted any MM charges less than four bonds removed from any QM atom. Waszkowycz et al. modified this approach so that only a single MM charge at the M_1 site was excluded [61].

The danger of these schemes is that simply deleting charges according to the connectivity will often result in the remaining MM atoms, as experienced by the QM region, appearing to have a net charge. Such an artefact will have a particularly serious effect on computed energies for processes in which the total charge of the QM region is modified, such as protonation reactions.

As discussed above, many force fields define atomic charges such that they form neutral charge groups. This feature can be used to minimise the artefacts generated when deleting a set of MM charges. In the L2 scheme [49] the charges on the whole charge group containing atom M_1 (Fig. 5) are neglected when building the QM Hamiltonian. While this will clearly remove some significant physical interactions, the fact that the charges removed will sum to zero will ensure that the total MM charge experienced by the QM calculation is correct. The leading term that is missing will be the dipole moment of the first charge group.

A series of tests on protonation reactions were used to evaluate the schemes L1 to L3 [49], using semiempirical, DFT and MP2 wavefunctions. The differences between the schemes were observed to be more pronounced for ab initio wavefunctions, as expected from an analysis of the influence of the integral approximations. L1 was recommended only at the semiempirical level; for ab initio wavefunctions the selective integral deletion was considered to lead

to problems of imbalance in the electrostatics for the link atom region. The L2 scheme was found to be robust for all types of wavefunctions, and for ab initio studies using force fields based on neutral groups it appears the obvious choice. Similar conclusions have been reached in a related work [62].

Charge shifting schemes. If deletion of charges is considered undesirable, perhaps because the neutral groups are so large that a significant part of the classical interaction would be lost, or because the force field is not defined in terms of neutral groups, an alternative approach is needed.

Consider the zeolite case shown in Fig. 4. Since the M_1 centres still have finite charge, further adjustments are necessary. The approach adopted sets M_1 to zero, but for each of the connected atoms M_2 a corresponding charge adjustment is made to conserve the total charge of the MM system. As M_1 and M_2 will usually have charges of opposite sign, the M_2 atoms will have reduced charges. To compensate for the dipole that has been created by the charge shift from M_1 to M_2 a pair of equal and opposite point charges are added close to each M_2 centre along the M_1–M_2 bond direction. The resulting MM charge distribution therefore has the same charge and dipole moment to that of the uncorrected MM 'defect', but the charge distribution close to the link atoms has been eliminated (Fig. 6).

Gaussian blur. Brooks has suggested [22] that the problems of close approach of the MM charges to the QM region are largely the result of the representation

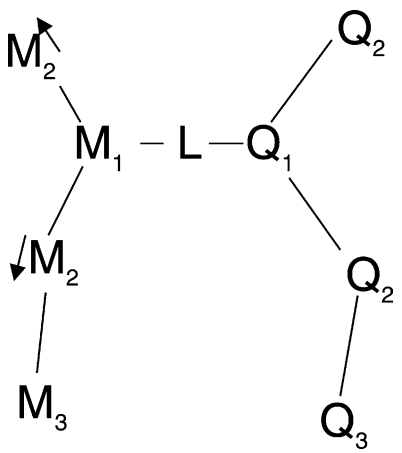

Fig. 6. Boundary charge shift approach.

of this charge distribution by a point charge model. The 'Gaussian blur' approach replaces the point charges for selected MM centres with a Gaussian charge distribution. An implementation of this approach is included in the coupling of GAMESS-UK and CHARMM [22] and is available in the QUASI software when GAMESS-UK is chosen as the QM code.

2.3.7. Link atom positions and forces

If the link atom coordinates can be written as a function of the real atom coordinates it is possible to eliminate them from the set of coordinates used in geometry optimisation or molecular dynamics. Such an elimination is particularly desirable for molecular dynamics and the evaluation of vibrational frequencies, which would be modified by coupling to any independent link atom motions. Since the link atoms have non-zero forces it is necessary to add a term to the real atom forces to account for the changes in the link atom positions resulting from movement of the real atoms. In the IMOMM scheme of Maseras and Morokuma [35] this was implicitly performed by working in internal coordinates such that the same internal coordinates were used to position the link atoms and to define the corresponding real atom (M_1) coordinates. The internal coordinate force can then be obtained by adding the QM and MM contributions.

When working in pure Cartesian coordinates the same effect can be realised by using the chain rule to establish the contribution to the forces on the real atoms such as M_1

$$\frac{\partial E}{\partial x_{M_1}} = \frac{\partial E}{\partial x_{M_1}} + \frac{\partial E}{\partial x_L}\frac{\partial x_L}{\partial x_{M_1}} \tag{5}$$

where the derivative $(\partial x_L/\partial x_{M_1})$ is a 3×3 matrix describing the coupling of the link atom and real atom motions. For the case of a link atom placed at a fixed position along the QM/MM bond there will be a correction term for the atom at each end of the bond in question (M_1 and Q_1 in Fig. 5). This approach is similar to that used in the QM-Pot scheme of Sauer et al. [40].

2.3.8. Adjustments to the classical energy expression

A practical requirement for any QM/MM implementation is that the MM energy expression is

modified to remove terms involving only QM atoms, while retaining those terms involving both QM and MM atoms needed to implement the desired coupling scheme. In the QUASI implementation this is performed internally by the interfaces to the classical codes DL_POLY and GROMOS for the zeolite and enzyme models, respectively.

As noted in Section 2.3.2, the presence of link atoms requires some correction to the total energy expression. Ideally the link atom contribution to the energy could be modelled and removed. However, in practice this is not simple to do in the context of additive QM/MM methods, as it requires accurate parameterisation of the energy surface of the link atoms. It is easy to create a potential energy surface which deviates strongly from the correct one, especially if large distortions of the link atoms occur. If these deviations are in the direction of unphysically low energies geometry optimisation algorithms will become unstable.

Alternatively, an approximate link atom correction can be made by removing terms from the force field energy that are duplicated by the distortion energy associated with the link atoms. As an example consider the link atom L in Fig. 5. Distortion of the angle $Q_2-Q_1-M_1$ away from equilibrium will lead to quantum mechanical restoring forces arising from the movement of link atom L from its equilibrium position. As described in Section 2.3.7 this will lead to a force on atom M_1. However, since atom M_1 is molecular mechanical in nature there will also be an MM term (an angle bending term in a valence force field) $Q_2-Q_1-M_1$ which provides a restoring force for the same distortion. To avoid multiple counting this term is removed from the hybrid energy expression, as are torsion angles of the form $Q_3-Q_2-Q_1-M_1$.

2.3.9. Models for ionic materials

In this section we present the implementation of hybrid QM/MM techniques developed within QUASI to study reactions at the surfaces of ionic solids including reconstructed polar surfaces and interfaces.

Background. A number of alternative approaches have been developed and implemented in computer codes in order to treat point defects in ionic solids with ICECAP [63–65], GUESS [66–71] and AIMP [72–76] being of particular interest to us. Within the framework of the discussion in Section 2.3.1 these are boundary region methods. The external embedding potential is commonly approximated as the electrostatic potential of the crystal remainder complemented by local or semilocal pseudopotentials sited on boundary atoms around the embedded cluster. Parameters of such potentials can be obtained in a self-consistent manner or borrowed and readjusted from those of standard atomic quantum-chemical pseudopotentials. Our method follows approaches commonly used for treatment of point defects in the bulk of halide and oxide materials, discussed above, but also accounts for surface (or interface) specific electrostatic effects and morphology, which allow one to study chemisorption of charged and strongly dipolar species as well as inherent structural and electronic surface defects.

Representation of ionic materials. In polar solids, such as alkali metal and alkali earth halides and oxides, individual ions present a good example of structural elements (Section 2.3.5). The reference system can be represented exclusively by ionic structural elements whilst in the defective system the whole defect region is considered as a single structural element. The force field representation of the classical part of the system will typically be based on the rigid-ion, or Born, model [77,78] which provided the first accurate atomistic treatment and insight into thermal cohesive, elastic and dielectric properties of ionic materials. In the Born model, long-range electrostatic interactions between spherical ions yield the predominant contribution to the potential energy of nuclei, whilst the short-range pairwise interactions maintain the system stability preventing ions from collapsing onto each other. The shell model of Dick and Overhauser [18] further extends the Born model to include an electronic component of polarisation for ions in the local fields created by all other ions in the system and external sources. In the shell model the ion is comprised of two charged components, a spherical shell of charge, to model the valence electrons which contribute to the polarisability, and a point charge core. Simple MM-based methods using polarisable ions have traditionally been very successful in the accurate description of both perfect periodic and defective polar solids and surfaces [79–81]. The response of most of

the support to the presence of the adsorbate can thus be expected to be well reproduced by such methods.

The embedding potential. The total energy of the defective system can be decomposed into the self-energy contributions of all structural elements with the energy of their interaction expanded in a series over a number of interacting elements in an exact analogy with the semiclassical (shell model) representation. For the purposes of modelling, the properties of the structural element that comprises the cluster with a defect are not known a priori and therefore should be obtained ab initio by a QM method whilst the structural elements in the system remainder can be treated using interatomic potentials at an MM level of theory. The resulting partition of the energy can be written as

$$E^{\text{Total}} = E^{\text{on-site}} + E^{\text{2-body}} + E^{\text{3-body}} + E^{\text{4-body}} + \cdots \quad (6)$$

The on-site contribution constitutes the self-energy of the structural elements

$$E^{\text{on-site}} = E^{\text{cluster}} + \sum_i E_i^{\text{shell}} \quad (7)$$

that is reduced to the expectation value of the usual molecular many-electron Hamiltonian, $\hat{H}^{\text{QM}}$, with the corresponding cluster group function, $|\Psi\rangle$, along with the energy of electrostatic interactions between nuclei in the QM cluster, U_{ab}^{nucl} :

$$E^{\text{cluster}} = \langle\Psi|\hat{H}^{\text{QM}}|\Psi\rangle + \frac{1}{2}{\sum_{a,b}}' U_{ab}^{\text{nucl}} \quad (8)$$

and the shell polarisation for all species treated at an MM level using a spring interatomic potential with harmonic and quartic terms:

$$E_i^{\text{shell}} = \frac{1}{2!}k_i^{(2)}d_i^2 + \frac{1}{4!}k_i^{(4)}d_i^4, \quad (9)$$

where $d_i = |\mathbf{r}_i^{\text{shell}} - \mathbf{r}_i^{\text{core}}|$.

(Throughout indices $i,j,\ldots$ are reserved for structural elements and $a,b,\ldots$ for nuclei in the cluster.)

In this work we neglect higher order many-body contributions to the total energy and will concentrate on pairwise interactions:

$$E^{\text{2-body}} = \sum_j U_{\text{cluster},j} + \frac{1}{2}{\sum_{i,j}}' U_{ij}, \quad (10)$$

with the embedding potential experienced by electrons in the cluster, along with the nuclei, expanded as a sum of contributions from all structural elements of the system remainder

$$U_{\text{cluster},j} = \langle\Psi|\hat{V}_j^{\text{short}} + \hat{V}_j^{\text{long}}|\Psi\rangle + \sum_a U_{aj}^{\text{nucl}}, \quad (11)$$

and all interactions in the system remainder accounted for using interatomic potentials

$$U_{ij} = U_{ij}^{\text{short}} + U_{ij}^{\text{long}}. \quad (12)$$

The short-range contribution from structural element j to the QM embedding potential, $\hat{V}_j^{\text{short}}$, can be approximated by the Hay and Wadt semilocal pseudopotentials [82–84]

$$\hat{V} = \hat{V}^{\text{local}} + \sum_{lm} w_l|lm\rangle\langle lm| \quad (13)$$

or non-local pseudopotentials due to Durand and Berthelat [85]

$$\hat{V} = \sum_{li} |\phi_{li}\rangle\epsilon_{li}\langle\phi_{li}| \quad (14)$$

where w_l and ϕ_{li} coefficients are represented using products of angular functions with polynomial and Gaussian radial functions, and ε_{li} are numerical coefficients.

Potentials (13) or (14) provide conditions for localisation of electrons in the cluster region preventing spillage of electrons to positive centres in the system remainder. They correct for the cluster terminating effects on the one-electron spectrum (particularly the narrowing of the gap between the highest occupied and lowest unoccupied states) and compensate for missing short-range forces in the interface between the QM cluster and the system remainder. Usual atomic, large-core pseudopotentials (e.g. LANL [82–84], Stuttgart-Dresden [86], etc.) prove to be a good approximation for cationic species. Anions, however, present a problem not only due to the lack of parameters available from the literature, but also in that most of the current molecular QM codes do not support such objects. This problem has been recently addressed in the GAMESS-UK code and work to include O^{2-} pseudopotentials will be reported in the near future. In the present study we have introduced a simplification and replaced the short-range QM contribution to the interaction energy

between the QM cluster and anion j by interatomic potentials

$$U_{\text{cluster},j} = \langle \Psi | \hat{V}_j^{\text{long}} | \Psi \rangle + \sum_a (U_{aj}^{\text{nucl}} + U_{aj}^{\text{short}}) \qquad (15)$$

retaining expression (11) for the cations only. Furthermore, as core electrons in the electron group $|\psi\rangle$ are often excluded from consideration, using atomic core pseudopotentials, short-range interactions between the cores of respective atoms disappear from E^{cluster} and $U_{\text{cluster},j}$ and should be compensated for. This problem also occurs in conventional quantum chemical calculations but is normally dealt with by including outer (core) electrons in the valence electron group. The remaining hard cores are effectively screened by the semicore electrons yielding negligibly small contributions to the total energy. In contrast, the large core pseudopotentials used for embedding necessitate inclusion of short-range classical interactions between cationic species in the QM cluster and the nearest neighbour cations.

The pairwise short-range interactions are modelled by parametrised analytical functions of the interionic distance. The shell model for polarisable ions postulates that the short-range interatomic forces operate between shells rather than the cores usually associated with nuclei. The model therefore includes the dependence of the short range interactions not only on the interatomic distance but also on the induced dipoles. When trying to find an exact match between the shell model and a QM description of an ion, an open question remains whether the short-range QM potential in Eq. (11) should be centred on the core or a shell of the corresponding ion. Although this problem disappears as cations are often considered to be rigid ions, for more polarisable species (cations from the lower part of the periodic table as well as anions) we anticipate the use of such potentials first of all on the cores, which is computationally less expensive in our scheme.

The long-range interactions that enter Eqs. (11), (12) and (15) are purely electrostatic for which we consider the far and the near field cases. The generic Coulomb potential due to a charge q_j far away from the cluster can be conveniently expanded as a Taylor series about the equilibrium position of an ionic centre in the reference system. Only the leading (monopole) term of the expansion has the long-range character and therefore requires special attention when considering the embedding potential as described in Section 2.3.3; the rest result in long-range polarisation effects included a posteriori. At closer distances to the defect centre, the displacements of atoms due to a defect can be large and therefore both Coulomb and short-range interactions are summed up explicitly within the active and frozen MM regions.

Finally, the mutual self-consistency in electronic polarisation effects between QM and MM regions is achieved by a series of microiterations. In particular the QM calculations in the embedding potential are followed by shell relaxations in the polarising potential due to the charge distribution in the QM region at the previous iteration. The coordinates of the nuclei in the QM region and cores from the active MM region are considered together to form the optimisation space while the group wave function along with the positions of shells adjust adiabatically.

2.4. The parallel implementation of the QUASI software

ChemShell is implemented in parallel using either MPI, with or without linking to the Global Array tools [87]. For large parallel machines the QM and MM modules required are linked into the same executable as the Tcl interpreter, so that execution of the whole QM/MM calculation can proceed without spawning additional processes. DL_POLY, GAMESS-UK and GULP are all available in this mode of operation. The Tcl interpreter only executes on node 0 (the master), but on each call to the parallel compute modules all nodes are deployed, typically for a single energy and gradient evaluation, before control is handed back to the interpreter. We have used parallel platforms (both supercomputer and commodity systems) extensively within the QUASI applications described below.

An example (thrombin in water) is shown in Fig. 7. In this example the QM region (69 atoms) is modelled by SCF 6-31G calculation (401 GTOs). The total system size for the classical calculation is 16,659 atoms. The MM calculation was performed using all non-bonded interactions (i.e. without a pairlist), and the QM/MM interaction was cut off at 15 Å using a neutral group scheme (leading to the inclusion in

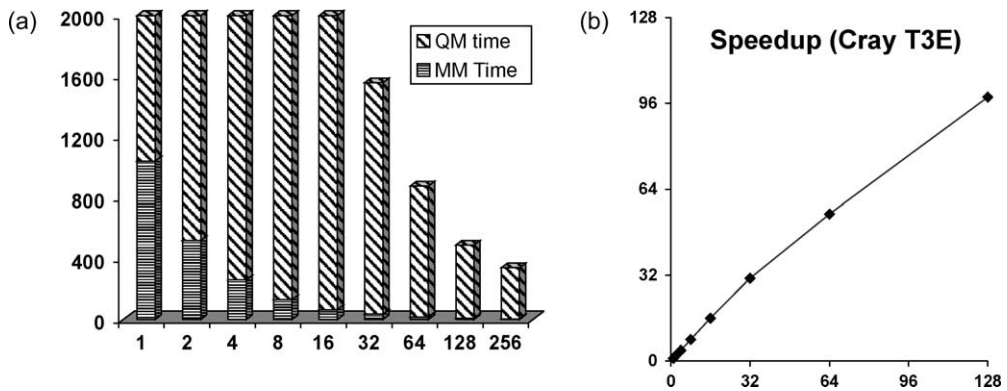

Fig. 7. Performance of a thrombin QM/MM calculation on the Cray T3E. (a) Time in seconds (cropped at 2000 s, the one processor time is 49,043 s). (b) Speedup.

the QM calculation of 1507 classical centres). Under these conditions the QM cost dominates (as is usually the case for our QM/MM calculations) but the scalability of the MM component is also apparent.

3. Some industrial applications of the QUASI software

In the following sections we describe some of the applications work performed within the QUASI project. The examples are chosen to illustrate the range of applicability of the general purpose QM/MM implementation developed within the project.

Throughout, basis sets and density functionals are as defined in the respective quantum chemistry programs, unless otherwise stated.

3.1. Applications to surface catalysis—adsorption of formate on the polar oxygen terminated surface of ZnO

The calculations in this section were performed at the Royal Institution in close collaboration with the research groups at ICI and Synetix.

3.1.1. Industrial relevance

Formate is the most stable of the intermediates in the industrial synthesis of methanol from a feed gas of $H_2/CO/CO_2$, over a Cu/ZnO catalyst [88]. Production by the low-pressure methanol technique using the Cu/ZnO catalyst of ICI/Synetix accounts for 60% of the world's methanol production. Although this is not new technology the mechanism by which methanol is produced is still open to debate as there is a paucity of mechanistic detail for environments that approach realistic industrial conditions.

Methanol synthesis over the ZnO catalyst proceeds through successive hydrogenation of CO_2 via formate as in the case of the Cu/ZnO industrial catalyst. We have used ZnO as a model for the working catalyst in these studies. Experimental characterisation of the formate ion is possible using combined infrared spectroscopy and temperature programmed desorption studies [89]. In particular, the formate signal was identified in the 1550–1650 cm^{-1} region, which is the dominant feature in the spectra, as well as at 1350–1400 cm^{-1}. Both uni- and bidentate adsorption scenarios have been considered in the literature with peaks at 1577, 1583, 1589, 1597 and 1610 cm^{-1} variously attributed to formate.

We have obtained the structure and vibrational modes for the unidentate adsorption mode of formate on the polar (000-1) surface of ZnO, an active surface for catalysis, using the solid state scheme introduced in Section 2.3.9.

3.1.2. Computational approach

The preparation of an embedded cluster is a multi-step procedure, including the generation of a finite cluster model from the perfect, periodic system (Section 2.3.3). We use as an example a model of an active catalytic site on the polar oxygen terminated surface of ZnO, which has been utilised in our applied studies of the methanol synthesis [90] and particularly of the formate intermediate discussed in detail below.

A two-dimensional periodic slab model of the two complementary reconstructed polar surfaces was fully relaxed using the MM methodology employing the computer code MARVIN [91]. To define our structural model a hemispherical cut of the surface was made at a 25 Å radius from the vacant oxygen interstitial surface site. The atoms in the central core of this hemisphere (radius 13 Å, 604 atoms) were allowed to relax fully in all calculations with the remaining peripheral atoms (4394 cores and shells) fixed at their optimised positions from the MARVIN procedure. The thickness of the peripheral region was fixed at the cut-off radius of the short range MM potentials of 12 Å. ChemShell was used to calculate the electrostatic potential due to the ZnO remainder throughout the core region, which was reproduced by a set of 269 point charges positioned around the hemisphere to within 10^{-4} V. The active site is modelled as an embedded 12-atom QM cluster with a surrounding interface of 13 zinc based pseudo-potentials, and classical oxygen anions that corresponds to a 3 Å cut-off from the QM atoms. The classical representation of the nearest neighbour interaction between the QM region and the oxygen sub-lattice was found to be robust and effective in this study.

We employed DFT, specifically the B97-1 exchange correlation functional and a triple zeta valence basis set with two polarisation functions (tzv2p), which has been optimised for the support and the adsorbate [92,93]. For the purpose of the current investigation we have tested a basis denoted AUG-CC, which is the aug-cc-pV5z basis [94,95] but with the f, g, and h functions removed. In our preliminary tests, although this has strongly affected total calculated correlation energies with the coupled cluster approach, we saw little effect on equilibrium geometries or harmonic frequencies. An intermediate quality aug-tzv2p basis set was defined by augmenting the tzv2p basis with diffuse functions (exponents scaled by a factor 0.333 for the most diffuse functions of each orbital moment).

Full geometry optimisation has been applied to all centres within a 13 Å radius of the active site. A two-point displacement method with analytical calculation of gradients has then been used to build a dynamic matrix. Vibrational modes were obtained by diagonalisation of the dynamic matrix for a specified number of centres, the number of centres used being varied (Table 2). Electronic (fast) polarisation of the environment is included in the frequency calculation, following the adiabatic approximation via complete self-consistent shell relaxation in the active MM region.

Table 2
Vibrational frequencies for chemisorbed formate

Number of centres	Vibrational frequencies (cm^{-1})					
	$\nu(A_1)$	$\nu(B_1)$	$\nu(A_1)$	$\nu(B_2)$	$\nu(B_2)$	$\nu(A_1)$
HCO_2^-	791	1072	1212	1356	1695	3086
HCO_2^- + 9 atoms	793	1072	1213	1356	1695	3086
HCO_2^- + 12 atoms	793	1073	1214	1356	1695	3086

3.1.3. Properties of the formate ion absorbed on ZnO

Adsorption of formate on the ZnO surface leads to polarisation of the molecule with the lower bond C–O_a being 1.33 Å compared with 1.23 Å for C–O_b (as shown in Fig. 8) and 1.25 Å for the gas phase anion molecule. Stabilisation of the negatively charged molecular ion by the surface leads to the symmetry being broken. O_a interacts strongly with two under coordinated Zn ions (2.09 and 2.17 Å) while the third O_a–Zn interaction is 2.29 Å. The reason for the asymmetry is that O_b interacts with the third Zn (2.14 Å) further stabilising the adsorbed molecule.

For the lower B_2 stretching frequency we achieve a very good agreement with experiment with our calculated value of 1356 cm^{-1} comparing well with 1363 cm^{-1}. However, our calculated value for the higher B_2 stretching frequency, 1695 cm^{-1} (which is the experimental signature

Table 3
Vibrational frequencies for gas phase formate

Method	Vibrational frequencies (cm^{-1})					
	$\nu(A_1)$	$\nu(B_1)$	$\nu(A_1)$	$\nu(B_2)$	$\nu(B_2)$	$\nu(A_1)$
B97-1/tzv2p	738	1037	1323	1434	1654	2604
B3LYP/tzv2p	730	1045	1317	1368	1641	2617
B3LYP/aug-tzv2p	734	1048	1317	1369	1634	2618
B3LYP/AUG-CC	745	1043	1336	1375	1644	2573
CCSD/tzv2p	723	1081	1312	1380	1629	2725
CCSD/AUG-CC	762	1074	1366	1414	1673	2688

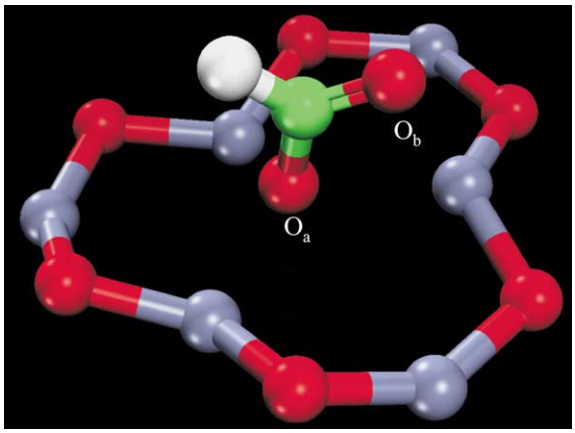

Fig. 8. The formate ion adsorbate and the portion of the surface that is included in the QM region (red: oxygen, green: carbon, white: hydrogen, grey: zinc).

peak for formate), differs by ca. 100 cm^{-1} from the experimental value; to investigate the causes of the discrepancy further calculations were performed increasing the numbers of centres involved in the vibrational modes. The results are summarised in Table 2, where mode assignments are based on the gas phase calculation of the formate ion.

The results indicate that the vibrational modes are not sensitive to the inclusion of surface centres. Visualisation of the key modes shows little involvement of the surface and therefore including only the adsorbate provides a reasonable comparison to experiment with a consequential saving in computational resources.

Further analysis has been performed on gas phase molecules in order to validate the exchange-correlation functional and basis set used Table 3. By comparing the vibrational modes calculated by B3LYP at the three basis set levels tabulated above, it is clear that the tzv2p basis reproduces satisfactorily the more extensive, high quality AUG-CC results; the CCSD calculations for gas phase formate employing AUG-CC, used here as a reference point, are in good agreement with our B97-1 tzv2p gas phase calculations, with the exception of the high frequency mode $\nu(A_1)$.

Therefore, we conclude that the current level of QM treatment in our hybrid calculations is sufficient, and we do not need to extend the number of sites in the harmonic frequency calculations. The difference with experiment is therefore assumed to be due to strong anharmonic effects in the ca. 1600 cm^{-1} asymmetric deformation mode and the size of the QM cluster (affecting the electronic structure of the anion adsorbate); nevertheless the accuracy of the harmonic approximation is in reasonable agreement with experiment. The implementation of techniques accounting for anharmonicity will be integrated into the ChemShell environment in the near future. Further application studies based on larger QM clusters are underway.

3.2. Application to zeolite catalysis

The zeolite application calculations within the QUASI project were carried out by Norsk Hydro, in collaboration with Daresbury Laboratory.

3.2.1. Industrial relevance

Nitrous oxide (N_2O) is an environmental pollutant, contributing to the catalytic destruction of stratospheric ozone, and is a greenhouse gas given its ability to absorb IR radiation. At present no legislation exists with regard to N_2O emissions. However, due to growing governmental awareness of the environmental impact of N_2O (United Nations Framework Protocol on Climate Change, Kyoto, Japan 1997), emission levels are expected to become regulated in the near future. At Norsk Hydro N_2O is formed as a by-product during production of nitric acid. Several materials (metal oxides, pervoskites and zeolites) show activity for N_2O decomposition, but so far the reaction rate is still too low to have any industrial significance. The lack of detailed understanding is often a problem for a more rational development or design of new and/or improved catalysts. A review of this subject is given by Kapteijn et al. [96].

The applications work has concentrated on Cu/ZSM-5. Experimental work by Li and Armor [97] shows that Cu exchanged ZSM-5 is an effective catalyst (as compared to a variety of other Co- and Cu-exchanged zeolites like beta, mordenite, ZSM-11 and ferrierite) for the decomposition of N_2O into its elements at 350–400 °C. The temperature regime reported in these works is preferable for the after treatment process of emission reduction.

The outline of this work is as follows; we first report on a comparison between the embedded cluster

and band structure calculations for Cu adsorption in chabazite, and for oxygen adsorption on the Cu site. In the second section we discuss binding of Cu at three different adsorption sites in ZSM-5. The results of this work are discussed in the light of previous results by Sauer and co-workers, where they have used a different QM/MM scheme. The last section is devoted to N_2O and NO adsorption on these three different Cu sites in ZSM-5. The catalytic cycle for N_2O decomposition on these Cu sites is also included.

3.2.2. Computational details

The work described here was performed with finite clusters, generated as described in Section 2.3.3, of typically 1100–1300 classical atoms. The valence force field of Hill and Sauer was used [14,15] with the charge shift correction (Section 2.3.6) at the QM/MM boundary. QM clusters contained typically 5 or more tetrahedral atoms (Si or Al) as indicated below. In the early stages of the project, effort was focussed on validating the QM/MM approaches by comparing different cluster sizes, and by testing against periodic DFT results. For this purpose we used the BP86 functional for comparison with periodic DFT results, such as those described below from VASP. However, as a precursor to the embedded cluster study of N_xO_y species a number of gas-phase DFT test calculations were carried out. These tests revealed that the hybrid B3LYP functional performed significantly better, in terms of the relative energetics, than BP86 so it was used for the studies reported in Sections 3.2.4 and 3.2.5.

3.2.3. Band structure vs. QM/MM: chabazite case study

For the validation of the QM/MM method we have performed band structure calculations. Experimental results, especially on the location of Cu within the zeolite are very uncertain and therefore not suitable as a reference. We have compared results using the T5 gas-phase cluster (Fig. 9), the embedded T5 cluster for chabazite, and chabazite bulk.

The reason for choosing chabazite is that the unit cell of ZSM-5 contains 288 atoms, so that detailed periodic DFT calculations were not feasible on the grounds of computational cost. The unit cell of chabazite contains only 36 atoms. However, we did not use chabazite for further investigation, such as adsorption studies, since the chabazite cavities are too narrow to host absorption on Cu of molecules such as NO.

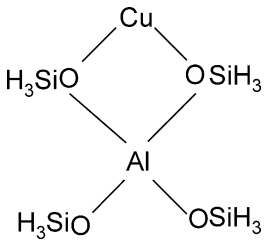

Fig. 9. The QM-cluster model (denoted as the T5 cluster) employed in this work.

For band structure calculations we have used the Vienna ab initio simulation program VASP [98–100]. The program is based on the density functional formalism and uses an iterative matrix-diagonalization scheme to solve the Kohn-Sham equations. Basis sets are plane waves combined with ultrasoft pseudopotentials; the generalized gradient approximation of Perdew-Wang-91 was used for non-local corrections. We used a cutoff of 495 eV and a set of 2k points in the Brillouin zone (Monkhorst-Pack division of $2 \times 2 \times 2$ which corresponds to a spacing of 0.05 $Å^{-1}$). The calculated lattice parameters for chabazite are in good agreement with experiment: $a = 9.38$ Å, $\alpha = 95°$ (exp. $a = 9.42$ Å, $\alpha = 94°$). The geometry of a T5 fragment of the bulk is presented in Table 4, together with results for gas phase and embedded systems.

Embedding improved the gas-phase cluster bond lengths. Cu–O bonds increase in the following order: 1.97, 1.99, 2.01, 2.05 Å for gas-phase, mechanical and

Table 4
Comparison of gas-phase T5, embedded T5 and bulk chabazite

Geometry (Å and degrees)	T_5^0	ZT_5^M	ZT_5^E	Band structure
Cu–O	1.97	1.95	1.97	2.05
	1.97	1.99	2.01	2.05
Cu–Al	2.62	2.65	2.67	2.78
Al–O	1.86	1.81	1.81	1.79
	1.86	1.84	1.82	1.80
O–Al–O	96	95	95	95
Si–O	1.69	1.68	1.70	1.63
	1.69	1.69	1.72	1.63

M and E denote mechanical and electrostatic embedding, respectively. Throughout the text we present geometry parameters related to oxygen atoms bonded to Cu.

electrostatic embedding, and band structure respectively. Similarly, Al–O bonds are improved by embedding: they decrease from 1.86 to 1.80 Å. The structural features resulting from electrostatic embedding calculations are closer to band structure results than those from mechanical embedding for Al–O and Cu–O bond lengths.

However, Si–O bonds are not improved by embedding, most likely because silicon atoms of the QM moiety, terminated by hydrogen atoms, are the junction atoms between QM and MM regions.

We have also compared the energetics of band structure and embedded calculations by studying the atomic oxygen adsorption. Here we have also carried out band structure calculations on the T5 cluster (using a supercell), including the adsorption of O, to provide an indication of the effects related only to the fundamental differences of methodologies (Becke-Perdew 86 vs. Perdew-Wang 91, localised basis set vs. plane waves, etc.) and thus how close we should expect QM/MM results to approach the band structure ones. The results are given in Table 5.

The oxygen adsorption energy is calculated to be 7.0 kcal/mol, which is in very good agreement with the VASP result of 5.4 kcal/mol, suggesting that band structure calculations of the type carried out here should be a reasonable reference for QM/MM calculations. Another point worth noticing is that Si–O bonds are still longer compared with VASP, even with embedding. This suggests that the deviation between the VASP and embedded results is only partially due to the description of the junction.

Table 5
Comparison of T5 cluster geometries and O adsorption energies as obtained from ChemShell and VASP (periodic super cell) calculations

Geometry (Å, degrees)	T_5^0 (Chemshell)	T_5^0 (VASP)
Cu–O	1.96, 1.97	1.97, 1.98
Cu–Al	2.80	2.80
Al–O	1.84	1.83
O–Al–O	88	88
Si–O	1.70, 1.71	1.65, 1.69
O–H	1.50	1.49
Cu–O_{ads}	1.72	1.72
ΔE_{ads} (kcal/mol)	−7.0	−5.4

The adsorption energy is calculated from the reaction $T_5 + (1/2)O_2 \rightarrow T_5$-O.

Table 6
Oxygen adsorption in Cu-exchanged Chabazite

Geometry (Å, degrees)	ZT_5^M-O	ZT_5^E-O	Band structure
Cu–O	2.00, 2.01	1.98, 2.02	2.02, 2.02
Cu–Al	2.69	2.70	2.79
Al–O	1.85, 1.86	1.80, 1.84	1.80, 1.80
O–Al–O	91	90	92
Si–O	1.69,1.69	1.69,1.72	1.66, 1.69
Cu–O_{ads}	1.72	1.72	1.73
ΔE_{ads} (kcal/mol)	+1.9	+1.6	−1.6

Comparison between mechanical embedding (M), electrostatic embedding (E) and band structure results.

The O adsorption results using T5 embedded in chabazite are given in Table 6 including the results of the band structure calculation. The adsorption energy decreases by about 9 kcal/mol upon introduction of embedding, as opposed to Cu/ZSM-5/NO where it increases by 10 kcal/mol. This could be due to larger Pauli repulsion between framework and adsorbed oxygen, given that the channels are narrower in chabazite. Both QM/MM adsorption energy and geometry are in good agreement with the band structure results. The electrostatic embedding results are somewhat closer to the band structure calculation, suggesting that the embedding describes the surroundings of the active site in an accurate fashion.

3.2.4. Coordination of Cu^+ Ions in Cu-ZSM5

Following the study of Nachtigallova et al. [42], we have considered three different active Cu sites for the adsorption study (Fig. 10). In all calculations we have included part of the MM region in

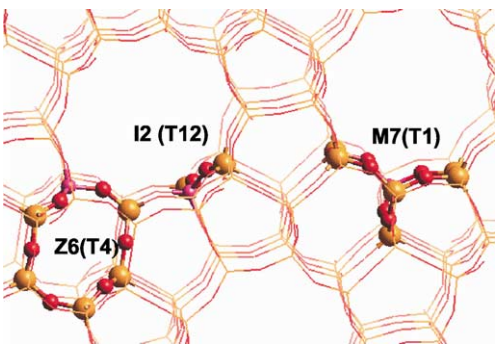

Fig. 10. Location of the sites in the zeolite framework. (Oxygen is given in red, silicon in yellow, aluminum in pink.).

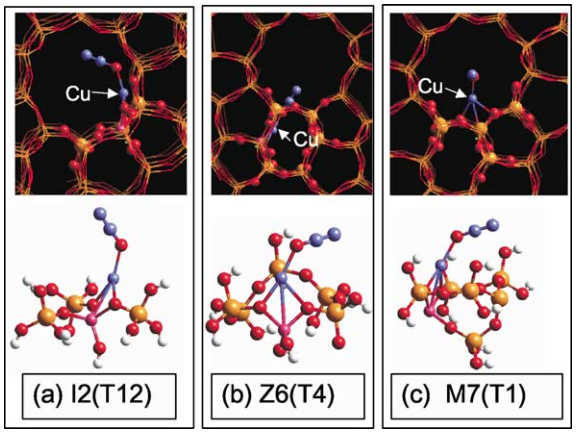

Fig. 11. Cu clusters, showing abdsorbed N_2O (a) I2(T12) (b) M7(T1) and (c) Z6(T4). Colours as for Fig. 1, with Cu and N in blue.

the optimisation. To define the active atoms for the optimisation we first identified all oxygen atoms with bonding contacts to Cu (there are two in the case of the I2 sites and four where Cu coordinates to a ring). The relaxed MM atoms are determined to be those four bonds from these oxygen atoms. This results in a number of active atoms ranging from 45 for the I2(T12) site, to 71 for the Z6(T4) site, and 68 for the M7(T1) site.

The I2(T12) site (Fig. 11(a)) is in the intersection of the main and the zig-zag (sinusoidal) channel. The Z6(T4) site is the most stable Cu configuration for the given Al substituted site. This site involves six Si atoms forming a ring structure (Fig. 11(b)). As for the M7 site, the T1 and T2 sites are among the most stable M7 sites and we have chosen a T1 Al site. The M7 site involves two fused ring structures, where a six Si atoms ring structure forms part of the wall of the main channel (Fig. 11(c)).

The most stable oxygen coordination sites are the same as found by Sauer et al. using a subtractive QM/MM scheme, denoted QM-Pot [42]. They obtain an energy difference between these two sites of about 3 kcal/mol, which can be compared with a value of 6.8 kcal/mol in the present study. The optimized Cu–O bond distances for those oxygen atoms coordinated to the Cu ion are displayed in Table 7, together with the corresponding QM-Pot results [42]. As Table 7 shows, there is a good correpondence between the different methodologies, in particular for the Cu–O(Al) bond distances in the I2 and Z6 systems.

Table 7
Cu–O distances at the I2, Z6 and M7 sites (Å)

Site	Chemshell	QM-Pot
I2(T12)	2.04, 2.05	2.04, 2.05
Z6(T4)	2.06, 2.13, *2.20*	2.08, *2.13*, 2.14
M7(T1)	2.11, 2.16, *2.21*	2.13, *2.20*, 2.24

The numbers in italics correspond to the Cu–O bond distances between Cu and oxygen atoms not adjacent to Al.

3.2.5. Geometries and energies for N_2O and its decomposition products

The most striking difference in the geometries obtained from bare cluster models and embedded cluster models is the non-linearity of the N_2O–Cu fragment in the embedded cases, as shown above in Fig. 11. A linear geometry has also been obtained by Schneider et al. using a bare $Cu(Al(OH)_4)$ cluster [101]. Since there are no close van der Waals contacts with the walls of the zeolite pores we believe the non-linearity arises from electrostatic interactions with the zeolite framework. Despite the geometrical difference, the adsorption energy is similar in both cases. For the I2(T12) site, the effects of the zeolite framework on adsorption energies are not very pronounced, that is, the difference for the oxygen end adsorption is only 1.1 and 6.4 kcal/mol for nitrogen end adsorption. The adsorption energies for N_2O and NO at the three Cu sites under study are given in Table 8. From the results there appears to be stronger adsorption on I2(T12) compared with the Z6 and M7 sites. However, the difference between the I2 and M7 sites is only on average about 2 kcal/mol. This small energy difference is reflected in the calculated geometry for the adsorbed species. The bond lengths of the coordinated N_2O are found to be similar on all three sites, with Cu–O and NO distances varying by no more than 0.02 Å.

Table 8
N_2O and NO adsorption energies on the I2, Z6 and M7 sites

System	Adsorption energy (kcal/mol)		
	I2 (T12)	Z6 (T4)	M7 (T1)
Z–Cu–NO	−30.2	−24.9	−28.0
Z–Cu–ON	−15.4	−10.7	−13.1
Z–Cu–NNO	−22.6	−19.3	−20.9
Z–Cu–ONN	−13.1	−11.2	−11.6

Table 9
Reaction energies for some key processes

Reaction	Reaction energy (kcal/mol)		
	I2 (T12)	Z6 (T4)	M7 (T1)
ZCu–NO + NO → ZCu + N_2 + O_2	−11.4	−16.5	−13.7
ZCu–ON + NO → ZCu–O + N_2O	−4.1	−4.8	−4.2
ZCu–NNO + NO → ZCu–NO + N_2O	−7.7	−5.9	−7.1
ZCu–ONN → ZCuO + N_2	8.4	10.1	9.1
ZCu–O + N_2O → ZCu + N_2 + O_2	−22.1	−25.8	−24.4
ZCu–O → ZCu + 1/2O_2	−8.7	−12.4	−11.0

For both ring sites M7 and Z6, the number of framework oxygen atoms coordinated to Cu is reduced from 3 to 2 upon adsorption of N_2O and NO, resulting in sites which resemble more closely an I2 site. Since both I2 and M7 are located in the main channel, this observation suggests that for these sites (Fig. 11), the catalytic activities are probably similar. For the Z6 site, the surroundings are different, suggesting that the transition state could be different compared to I2 and M7 sites.

Calculated reaction energies for some key reactions in the catalytic cycle are given in Table 9. Again, there are no pronounced differences between the three sites. For instance, the decomposition energies of N_2O (through O end adsorption) to N_2 and O_2 differ by only 1.5 kcal/mol. Similar results are also obtained using smaller QM clusters, suggesting that the computed energies are not highly sensitive to the computational conditions.

Although the systematic comparison of the energetics of N_2O decomposition at the Cu sites is still underway, we are able to summarise some of the results obtained to date by the reaction scheme shown in Fig. 12.

Many of these elementary reactions have been studied by Schneider et al. in a series of calculations on NO and N_2O decomposition [101,102] employing the bare Cu(Al(OH)$_4$) cluster. We see some significant quantitative differences arising from the use of smaller, bare clusters and a different level of DFT theory (LSDA). For example, the decomposition of N_2O (forming N_2 and 1/2O_2) is found to be thermoneutral at the LSDA level, in contrast to our

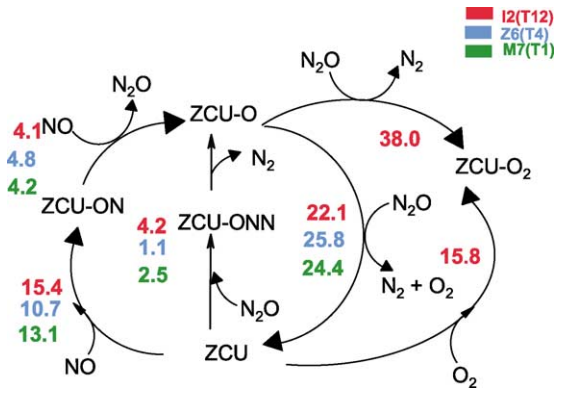

Fig. 12. Reaction energetics (kcal/mol) for N_2O decomposition on Cu–ZSM5.

B3LYP results which yield an exothermicity of 13.5 kcal/mol.

Our calculated reaction energies suggest that ZCuO is an important intermediate in the decomposition of N_2O to $N_2 + O_2$. Furthermore, the calculations also show that O_2 inhibits N_2O decomposition by forming a stable Z–Cu–O_2 species, which is in agreement with experimental observations [96]. Further studies, including the characterisation of the embedded transition states, are ongoing.

3.3. Modelling enzyme structure and reactivity

One of the major challenges in life science research is the detailed understanding of enzymatic reactions at the molecular level. QM methods can describe all biochemical processes, in principle, but they are still prohibitively expensive for systems as large as enzymes. The simpler MM methods are widely used for biomolecular simulations, but they cannot capture electronic events such as the making and breaking of bonds. Therefore, combined QM/MM methods offer a natural choice to model enzymatic reactions, since the active site and its protein environment can be treated at appropriate QM and MM levels, respectively.

The QM/MM models that have been implemented in the ChemShell software have recently been applied in several studies of enzymatic reactions. In this section we address three such applications. Since most of this work has already been published [20,30,103] we only summarize the main results in order to

indicate how the QM/MM functionality of ChemShell can be utilized in such research.

3.3.1. Enzymatic reactions in triosephosphate isomerase (TIM)

Systematic QM/MM calculations [30] were carried out for the TIM catalysed conversion of dihydroxyacetone phosphate into glycerinaldehyde 3-phosphate which proceeds by a four-step proton shuttle mechanism via an enediol. The primary objective was to investigate the sensitivity of optimized structures and relative energies towards variations in the QM/MM model, including the choice of the QM method, the size of the QM region, the size of the optimized MM region, and the treatment of the QM/MM boundary through various link atom and connection atom schemes. The QM methods that were applied in combination with the MSI-CHARMm force field range from semiempirical (AM1) to density functional (BP86, B3LYP) and ab initio (MP2) methods; the most extensive QM calculations involve 275 QM atoms and 2162 basis functions.

The main methodological conclusions from these comparisons are as follows: AM1/CHARMm and DFT/CHARMm predict qualitatively the same mechanism, with reasonable agreement in the computed relative energies and geometries. In most cases, AM1/CHARMm and DFT/CHARMm find the same hydrogen bonding pattern in the active site (except for one intermediate). The QM/CHARMm results with small and large QM regions (37 vs. 275 atoms) tend to be more uniform for BP86 than for AM1. The transition to a larger QM region preserves the basic energy profile, but causes considerable shifts in relative energies (typically ca. 3 kcal/mol). In the case of the larger QM region, the energy profile is determined by the QM contributions, while the MM contributions are non-negligible when using the small QM region. In AM1/CHARMm, different link atom and connection atom schemes yield very similar results (energies within 1 kcal/mol) (Fig. 13).

From a mechanistic point of view, the results support the four-step proton transfer pathway via an enediol, with involvement of His95 acting as a proton donor. The conversion between the enediolates with participation of His95 contains the rate-limiting step and proceeds via an intermediate enediol which is only a very shallow minimum. The alternative three-step

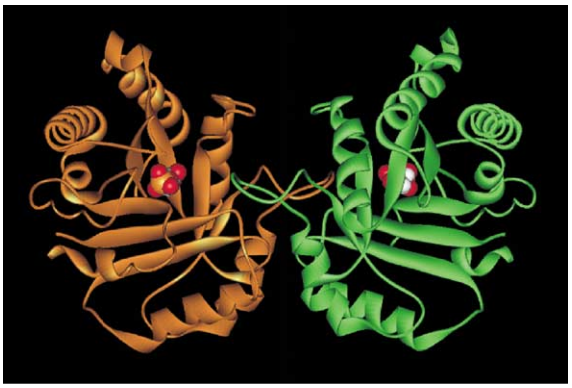

Fig. 13. Triosphophate isomerase dimer: schematic representation showing the substrate binding sites.

mechanism, without involvement of His95 and with a direct intramolecular proton transfer in the enediolate, requires more activation in the enzyme since it is disfavoured by the protein environment. Similar mechanistic conclusions have been reached in an independent B3LYP/CHARMM study which considered the two mechanisms outlined above and a third pathway [104,105].

3.3.2. Enzymatic reactions in p-hydroxybenzoate hydroxylase (PHBH)

PHBH is a flavoprotein, involved in the degradation of aromatic compounds, which catalyses the monooxygenation of *p*-hydroxybenzoate to form 3,4-dihydroxybenzoate. The mechanisms of the hydroxylation reactions in PHBH were studied at the QM/MM level both by molecular dynamics (MD) simulations and by geometry optimisation techniques [20]. The ChemShell code was used to locate the relevant minima and transition states for large model systems (AM1) and for the complete enzyme (AM1/GROMOS). The latter optimisations involved around 7000 atoms (including 102 QM atoms) and could not have been accomplished without the HDLC optimiser.

The molecular dynamics simulations (AM1/GROMOS) confirm that the dianion mechanism (involving deprotonated *p*-hydroxybenzoate) is favoured by a wide margin over the alternative monoanion mechanism (without deprotonation of the substrate), in agreement with previous QM/MM reaction path calculations [106]. The computed free energy barrier of 12 kcal/mol from these simulations is close to the available experimental estimate [107]. Geometry

Fig. 14. *p*-Hydroxybenzoate hydroxylase: schematic representation showing the substrate (red), cofactor (blue) and unrestrained active-site environment (green).

optimisations of gas-phase model systems of increasing size indicate the structural and electronic features that influence the activation energy. Compared with the largest model system (substrate + cofactor, 102 atoms), the initially located transition state in the full enzyme has a qualitatively similar structure and a somewhat lower activation energy, but the direct energetic influence of the protein environment on the barrier appears to be rather small [20] (Fig. 14).

In an extension of this work, several other representative transition states for this reaction in PHBH have been located at the AM1/GROMOS level, which differ to some extent in the protein conformation adopted, but the computed activation energies still fall into a rather narrow range [108].

3.3.3. The oxidant species of cytochrome P450 enzymes

Cytochrome P450 enzymes constitute a superfamily of monooxygenases that appear in all bioorganisms and perform vital bioregulatory functions. Large parts of their catalytic cycle are well understood, but the primary oxidant of cytochrome P450 (Compound I) is hard to detect experimentally. In the case of cytochrome P450cam (with the natural substrate camphor), this intermediate does not accumulate in solution during the reaction even at temperatures as low as 200 K [109]. This elusive species is thus a rewarding target for theoretical characterisation. The system studied [103] at the B3LYP/CHARMM level comprised the heme protein with the camphor substrate surrounded by a water layer of 16 Å thickness (24,394 atoms including 16,965 solvent atoms). Starting from related X-ray data [110], initial structures were prepared by classical MD simulations, and several snapshots from the equilibration trajectory were then taken as inputs for QM/MM geometry optimisations (using QM regions of 40–57 atoms). The results from these QM/MM calculations were compared with those from corresponding QM calculations on the isolated QM region, in order to analyse the effects of the protein environment (Fig. 15).

Compound I is an open-shell species with three unpaired electrons, two of which are localised in the FeO moiety while the third one can be delocalised over the porphyrin ring and the sulfur atom of the cysteinate ligand. The B3LYP/CHARMM calculations show that Compound I is transformed by the protein environment from a sulfur-centered radical to a 'green' porphyrin-centered radical cation (carrying ca. 70% of the relevant spin density). The hydrogen-bonding machinery of the protein pocket stabilizes the negative charge at the sulfur atom, and it shortens and strengthens the Fe–S bond. Calculations for different conformations of the protein pocket indicate that the balance of the sulfur/porphyrin radical character and the precise Fe–S bond length are finely tuned by small changes of the pocket. Compound I thus behaves as a chameleon species that can adapt its electronic and structural character to the specific environment.

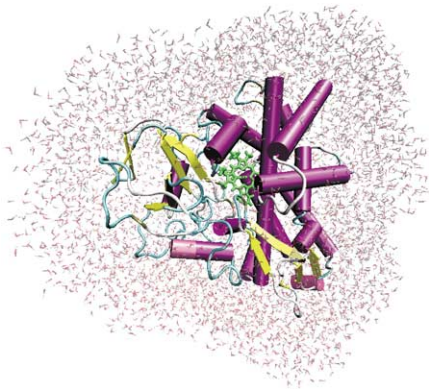

Fig. 15. Cytochrome P450: schematic representation showing the active-site porphyrin (green) and the surrounding water layer.

4. Summary

We believe that the QUASI project has demonstrated that QM/MM methods have developed to the point that they now have wide applicability to industrial modelling problems. We have implemented the approach within a general purpose framework that permits the use of a range of coupling methods to be tested on a given problem, and which allows the QM and MM codes to be chosen on the basis of applicability to the problem under study. The calculations are computationally demanding and we have therefore designed the software to exploit modern parallel computer systems, including both high-end and commodity class architectures. For more information on availability of the software described in this article please see the web site http://www.cse.clrc.ac.uk/qcg/quasi.

Acknowledgements

We thank the EU for financial support under project EP25047. We have had the benefit of consistent encouragement from our project officers in Brussels (Pertti Jauhiainen and Georgia Efthymiopolou) and review team (Erkki Brandas, Sean Corish, Bernard Maigret and Jose Brokken). We would also like to thank all others in the academic and industrial partner organisations who have helped with the smooth running of the collaboration.

References

[1] EUROPORT: see http://www.gmd.de/SCAI/europort-1/EUROPORT.HTM
[2] A. Warshel, M. Levitt, J. Mol. Biol. 103 (1976) 227.
[3] J. Aqvist, A. Warshel, Chem. Rev. 93 (1993) 2523–2544.
[4] J. Gao, in: K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 7, VCH, New York, 1995, p. 119.
[5] T.Z. Mordasini, W. Thiel, Chimia 52 (1998) 288–291.
[6] P. Sherwood, in: J. Grotendorst (Ed.), Modern Methods and Algorithms in Quantum Chemistry, NIC Series, Jülich, 2000, pp. 257–277.
[7] GAMESS-UK is a package of ab initio programs written by M.F. Guest, J.H. van Lenthe, J. Kendrick, K. Schöffel, P. Sherwood, with contributions from R.D. Amos, R.J. Buenker, H.J.J. van Dam, M. Dupuis, N.C. Handy, I.H. Hillier, P.J. Knowles, V. Bonacic-Koutecky, W. von Niessen, R.J. Harrison, A.P. Rendell, V.R. Saunders, A.J. Stone, D.J. Tozer, A.H. de Vries. The package is derived from the original GAMESS code due to M. Dupuis, D. Spangler, J. Wendoloski, NRCC Software Catalog, vol. 1, Program No. QG01 (GAMESS), 1980. See http://www.dl.ac.uk/CFS
[8] W. Thiel, MNDO99. Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim, Germany, 1999
[9] P. Sherwood, A. H.de. Vries, ChemShell User Manual, 1997–2002, available online at http://www.cse.clrc.ac.uk/qcg/chemshell
[10] J.K. Ousterhout, Tcl and the Tk toolkit, Addison Wesley, Reading, MA, 1994.
[11] W.L. DeLano, The PyMOL Molecular Graphics System (2002) on World Wide Web http://www.pymol.org
[12] W. Smith, T. Forester, J. Mol. Graph. 14 (1996) 136.
[13] W. Smith, C.W. Yong, P.M. Rodger, Mol. Sim. 28 (2002) 385–471.
[14] J.R. Hill, J. Sauer, J. Phys. Chem. 99 (1995) 9536–9550.
[15] J.R. Hill, J. Sauer, J. Phys. Chem. 98 (1994) 1238–1244.
[16] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, J. Comput. Chem. 4 (1983) 187–217.
[17] J.D. Gale, J. Chem. Soc., Faraday Trans. 93 (1997) 629–637.
[18] B.G. Dick, A.W. Overhauser, Phys. Rev. 112 (1958) 90.
[19] W.F.v. Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Krüger, A.E. Mark, W.R.P. Scott, I.G. Tironi, Biomolecular Simulation: The GROMOS96 Manual and User Guide, Biomos, Zürich, Groningen, 1996.
[20] S.R. Billeter, C.F.W. Hanser, T.Z. Mordasini, M. Scholten, W. Thiel, W.F. van Gunsteren, Phys. Chem. Chem. Phys. 3 (2001) 688–695.
[21] S.P. Greatbanks, P. Sherwood, I.H. Hillier, J. Phys. Chem. 98 (1994) 8134–8139.
[22] D. Das, K.P. Eurenius, E.M. Billings, P. Sherwood, D.C. Chatfield, M. Hodošček, B.R. Brooks, J. Chem. Phys. 117 (2002) 10534–10547.
[23] P. Sherwood, A.H. de Vries, S.J. Collins, S.P. Greatbanks, N.A. Burton, M.A. Vincent, I.H. Hillier, Faraday Discuss. (1997) 79–92.
[24] W. Thiel, in: J. Grotendorst (Ed.), Modern Methods and Algorithms in Quantum Chemistry, NIC Series, Jülich, 2000, pp. 233–255.
[25] D. Bakowies, W. Thiel, J. Phys. Chem. 100 (1996) 10580–10594.
[26] I. Antes, W. Thiel, J. Phys. Chem. A 103 (1999) 9290–9295.
[27] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, Chem. Phys. Lett. 162 (1989) 165–169.
[28] M. von Arnim, R. Ahlrichs, J. Comput Chem. 19 (1998) 1746–1757.
[29] Details online at www.turbomole.de
[30] C. Lennartz, A. Schäfer, F. Terstegen, W. Thiel, J. Phys. Chem. B 106 (2002) 1758–1767.
[31] J. Baker, D. Kinghorn, P. Pulay, J. Chem. Phys. 110 (1999) 4986–4991.

[32] K. Nemeth, O. Coulaud, G. Monard, J.G. Angyan, J. Chem. Phys. 114 (2001) 9747–9753.
[33] S.R. Billeter, A.J. Turner, W. Thiel, Phys. Chem. Chem. Phys. 2 (2000) 2177–2186.
[34] J. Baker, A. Kessi, B. Delley, J. Chem. Phys. 105 (1996) 192–212.
[35] F. Maseras, K. Morokuma, J. Comput. Chem. 16 (1995) 1170–1179.
[36] A.J. Turner, V. Moliner, I.H. Williams, Phys. Chem. Chem. Phys. 1 (1999) 1323–1331.
[37] Y.K. Zhang, H.Y. Liu, W.T. Yang, J. Chem. Phys. 112 (2000) 3483–3492.
[38] M. Sierka, J. Sauer, J. Chem. Phys. 112 (2000) 6983–6996.
[39] M. Bühl, F. Terstegen, F. Löffler, B. Meynhardt, S. Kierse, M. Müller, C. Näther, U. Lüning, Eur. J. Org. Chem. (2001) 2151–2160.
[40] U. Eichler, C.M. Kölmel, J. Sauer, J. Comput. Chem. 18 (1997) 463–477.
[41] M. Sierka, J. Sauer, Faraday Discuss. (1997) 41–62.
[42] D. Nachtigallova, P. Nachtigall, M. Sierka, J. Sauer, Phys. Chem. Chem. Phys. 1 (1999) 2019–2026.
[43] U.C. Singh, P.A. Kollman, J. Comput. Chem. 7 (1986) 718–730.
[44] M.J. Field, P.A. Bash, M. Karplus, J. Comput. Chem. 11 (1990) 700–733.
[45] D.M. Philipp, R.A. Friesner, J. Comput. Chem. 20 (1999) 1468–1494.
[46] P.L. Cummins, J.E. Gready, Chem. Phys. Lett. 225 (1994) 11–17.
[47] D. Bakowies, W. Thiel, J. Comput. Chem. 17 (1996) 87–108.
[48] B.T. Thole, Chem. Phys. 59 (1981) 341–350.
[49] I. Antes, W. Thiel, in: J.L. Gao, M.A. Thompson (Eds.), Combined Quantum Mechanical and Molecular Mechanical Methods, ACS Symposium Series, vol. 712, 1998, p. 50.
[50] B.T. Thole, P.T.v. Duijnen, Theor. Chim. Acta 55 (1980) 307.
[51] M.A. Thompson, G.K. Schenter, J. Phys. Chem. 99 (1995) 6374–6386.
[52] M.A. Thompson, J. Phys. Chem. 100 (1996) 14492–14507.
[53] P.T. van Duijnen, F. Grozema, M. Swart, J. Mol. Struct. (THEOCHEM) 464 (1999) 191–198.
[54] P.E. Smith, W.F. van Gunsteren, in: P.K. Weiner, A.J. Wilkinson (Eds.), Computer Simulation of Biomolecular Systems, vol. 2, ESCOM, Leiden, 1993, pp. 182–212.
[55] L.N. Kantorovich, J. Phys. C 21 (1988) 5041–5056.
[56] L.N. Kantorovich, Int. J. Quantum Chem. 78 (2000) 306–330.
[57] L.N. Kantorovich, Int. J. Quantum Chem. 76 (2000) 511–534.
[58] R. McWeeny, Methods of Molecular Quantum Mechanics, 2nd ed., Academic Press, London, 1992.
[59] J.H. Lii, N.L. Allinger, J. Comput. Chem. 12 (1991) 186–199.
[60] I. Antes, PhD Thesis, University of Zürich, 1998
[61] B. Waszkowycz, I.H. Hillier, N. Gensmantel, D.W. Payling, J. Chem. Soc., Perkin Trans. 2 (1991) 1819–1832.
[62] N. Reuter, A. Dejaegere, B. Maigret, M. Karplus, J. Phys. Chem. A 104 (2000) 1720–1735.
[63] J.H. Harding, A.H. Harker, P.B. Keegstra, R. Pandey, J.M. Vail, C. Woodward, Physica B & C 131 (1985) 151–156.
[64] A.B. Kunz, J.M. Vail, Phys. Rev. B 38 (1988) 1058–1063.
[65] J.M. Vail, R. Pandey, A.B. Kunz, Rev. Solid State Sci. 5 (1991) 241.
[66] A.L. Shluger, E.A. Kotomin, L.N. Kantorovich, J. Phys. C 19 (1986) 4183–4199.
[67] A.L. Shluger, A.H. Harker, V.E. Puchin, N. Itoh, C.R.A. Catlow, Modell. Simul. Mater. Sci. Engng 1 (1993) 673–692.
[68] A.L. Shluger, J.D. Gale, Phys. Rev. B 54 (1996) 962–969.
[69] A.L. Shluger, P.V. Sushko, L.N. Kantorovich, Phys. Rev. B 59 (1999) 2417–2430.
[70] P.V. Sushko, A.L. Shluger, C.R.A. Catlow, Surf. Sci. 450 (2000) 153–170.
[71] J.S. Braithwaite, P.V. Sushko, K. Wright, C.R.A. Catlow, J. Chem. Phys. 116 (2002) 2628–2635.
[72] S. Huzinaga, L. Seijo, Z. Barandiaran, M. Klobukowski, J. Chem. Phys. 86 (1987) 2132–2145.
[73] L. Seijo, Z. Barandiaran, J. Math. Chem. 10 (1992) 41–56.
[74] M.A. Nygren, L.G.M. Pettersson, Z. Barandiaran, L. Seijo, J. Chem. Phys. 100 (1994) 2010–2018.
[75] L. Seijo, Z. Barandiaran, Int. J. Quantum Chem. 60 (1996) 617–634.
[76] J.L. Pascual, L. Seijo, J. Chem. Phys. 102 (1995) 5368–5376.
[77] M. Born, K. Huang, Dynamical Theory of Crystal Lattices, Oxford University Press, Oxford, 1954.
[78] G. Venkataraman, L.A. Feldkamp, V.C. Sahni, Dynamics of Perfect Crystals, MIT Press, Cambridge, 1975.
[79] T.J. Grey, J.D. Gale, D. Nicholson, Mol. Phys. 98 (2000) 1565–1573.
[80] M.S.D. Read, M.S. Islam, F. King, F.E. Hancock, J. Phys. Chem. B 103 (1999) 1558–1562.
[81] L. Whitmore, A.A. Sokol, C.R.A. Catlow, Surf. Sci. 498 (2002) 135–146.
[82] P.J. Hay, W.R. Wadt, J. Chem. Phys. 82 (1985) 270–283.
[83] P.J. Hay, W.R. Wadt, J. Chem. Phys. 82 (1985) 299–310.
[84] W.R. Wadt, P.J. Hay, J. Chem. Phys. 82 (1985) 284–298.
[85] P. Durand, J.C. Berthelat, Theor. Chim. Acta 38 (1975) 283.
[86] M. Dolg, U. Wedig, H. Stoll, H. Preuss, J. Chem. Phys. 86 (1987) 866–872.
[87] J. Nieplocha, R.J. Harrison, R.J. Littlefield, J. Supercomput. 10 (1996) 197–220.
[88] M. Bowker, Vacuum 33 (1983) 669–685.
[89] S. Bailey, G.F. Froment, J.W. Snoeck, K.C. Waugh, Catal. Lett. 30 (1995) 99–111.
[90] S.A. French, A.A. Sokol, S.T. Bromley, C.R.A. Catlow, S.C. Rogers, F. King, P. Sherwood, Angew. Chem.-Int. Edit. 40 (2001) 4437.
[91] D.H. Gay, A.L. Rohl, J. Chem. Soc., Faraday Trans. 91 (1995) 925–936.
[92] R. Krishnan, J.S. Binkley, R. Seeger, J.A. Pople, J. Chem. Phys. 72 (1980) 650.

[93] A.D. McLean, G.S. Chandler, J. Chem. Phys. 72 (1980) 5639.
[94] T.H. Dunning, J. Chem. Phys. 90 (1989) 1007–1023.
[95] R.A. Kendall, T.H. Dunning, R.J. Harrison, J. Chem. Phys. 96 (1992) 6796–6806.
[96] F. Kapteijn, J. RodriguezMirasol, J.A. Moulijn, Appl. Catal., B 9 (1996) 25–64.
[97] Y.J. Li, J.N. Armor, Appl. Catal., B 1 (1992) L21–L29.
[98] G. Kresse, J. Hafner, Phys. Rev., B 47 (1993) 558–561.
[99] G. Kresse, J. Hafner, Phys. Rev., B 49 (1994) 14251–14269.
[100] G. Kresse, J. Furthmuller, Comput. Mater. Sci. 6 (1996) 15–50.
[101] W.F. Schneider, K.C. Hass, R. Ramprasad, J.B. Adams, J. Phys. Chem., B 102 (1998) 3692–3705.
[102] W.F. Schneider, K.C. Hass, R. Ramprasad, J.B. Adams, J. Phys. Chem., B 101 (1997) 4353–4357.
[103] J.C. Schöneboom, H. Lin, N. Reuter, W. Thiel, S. Cohen, F. Ogliaro, S. Shaik, J. Am. Chem. Soc. 124 (2002) 8142–8151.
[104] Q. Cui, M. Karplus, J. Am. Chem. Soc. 123 (2001) 2284–2290.
[105] Q. Cui, M. Karplus, J. Phys. Chem., B 106 (2002) 1768–1798.
[106] L. Ridder, A.J. Mulholland, J. Vervoort, I. Rietjens, J. Am. Chem. Soc. 120 (1998) 7641–7642.
[107] W.J.H. van Berkel, F. Müller, Eur. J. Biochem. 179 (1989) 307–314.
[108] S. Thiel, W. Thiel, in preparation
[109] R. Davydov, T.M. Makris, V. Kofman, D.E. Werst, S.G. Sligar, B.M. Hoffman, J. Am. Chem. Soc. 123 (2001) 1403–1415.
[110] I. Schlichting, J. Berendzen, K. Chu, A.M. Stock, S.A. Maves, D.E. Benson, B.M. Sweet, D. Ringe, G.A. Petsko, S.G. Sligar, Science 287 (2000) 1615–1622.