DRUG DISCOVERY TODAY BIOSILICO

Modelling enzyme reaction mechanisms, specificity and catalysis

Adrian J. Mulholland

Modern modelling methods can now give uniquely detailed understanding of enzyme-catalyzed reactions, including the analysis of mechanisms and the identification of determinants of specificity and catalytic efficiency. A new field of computational enzymology has emerged that has the potential to contribute significantly to structure-based design and to develop predictive models of drug metabolism and, for example, of the effects of genetic polymorphisms. This review outlines important techniques in this area, including quantum-chemical model studies and combined quantum-mechanics and molecular-mechanics (QM/MM) methods. Some recent applications to enzymes of pharmacological interest are also covered, showing the types of problems that can be tackled and the insight they can give.

Techniques capable of modelling enzyme reactions and analysing determinants of enzyme activity and specificity have the potential to contribute significantly to drug design and development. Perhaps the most obvious benefits are in structure-based design, where mechanistic knowledge (e.g. identifying key catalytic interactions at an active site) should assist in rational ligand design. Less obvious but increasingly important is the prediction of drug breakdown or activation by enzymes [1], enzyme-mediated adverse reactions [2] and their potential dependence on genetic polymorphisms, as for example in cytochrome P450 [2–4] (Figure 1) and glutathione S-transferase [5]. Modelling methods for mechanisms can now address these crucial questions about enzyme-catalyzed reactions and make predictive and practical contributions in all these areas.

In drug discovery, enzymes are important first of all as targets (many drugs are enzyme inhibitors), and ligand design should significantly benefit from knowledge of their mechanisms. For example, modelling can show uniquely how transition states and intermediates are stabilized within enzymes (which often show exceptionally high affinity for these unstable species), therefore helping inhibitor design (Figure 2). Mechanistic modelling adds another dimension to structure-based ligand design for enzyme targets. Analysis of these biochemical mechanisms will also help develop better predictive models of protein targets (e.g. kinases). A better understanding of the enzyme systems involved is also needed for the reliable prediction of pharmaceutical metabolism and toxicology (ADME–TOX) properties. Developments in pharmacogenomics will require models to predict the effects of genetic variation on enzyme activity and specificity. Similarly, developments in systems biology will require better quantitative and predictive understanding of enzyme action. For example, reliable rates for biochemical reactions are required as input for modelling of cellular and metabolic networks. Reliable methods to predict the effects of mutations on these rates are needed. Enzymes are increasingly used to catalyze processes in pharmaceutical synthesis, and there is

Adrian J. Mulholland
Centre for Computational Chemistry,
School of Chemistry,
University of Bristol,
Bristol BS8 1TS,
UK
e-mail:
Adrian.Mulholland@bris.ac.uk

1359-6446/05/$ – see front matter PII: S1359-6446(05)03611-1

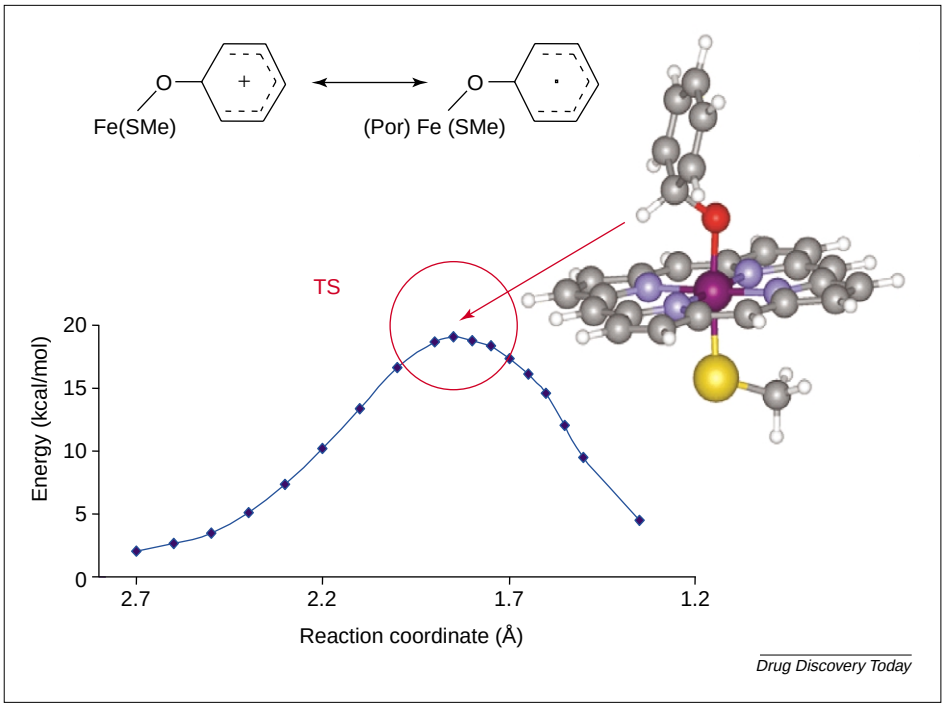

FIGURE 1

Aromatic hydroxylation in cytochrome P450, a key reaction in drug metabolism. Calculations on small models can identify mechanisms of enzyme-catalyzed reactions and aid in the development of SARs. The energy profile for aromatic hydroxylation of benzene by a model of the reactive P450 Compound I is shown (calculated with the B3LYP hybrid density-functional quantum-chemical method). Calculations allow the transition state (TS, the highest energy point along the reaction pathway) for the reaction to be found and analyzed: its structure is shown on the right (the iron atom form the porphyrin is coloured in blue, the oxygen atom in red and the sulphur atom in yellow). Two alternative modes of reaction were found, both of which might have relevance for drug metabolism [4]. Analysis of the transition-state electronic properties (e.g. its charge and spin distributions) shows it has mixed cationic and radical character (top left). This insight was crucial in developing a new, predictive, structure–reactivity relationship using simple descriptors [3,4].

a clear need to extend beyond the activities and specificities of naturally occurring enzymes for biocatalytic applications, for example with specifically designed protein catalysts. In all of these areas, molecular modelling of enzyme activity can make important practical contributions.

The field of computer modelling of enzyme reactions has matured to such a degree that it is now realistic and sensible to talk about computational enzymology [6,7]. Computer modelling and simulation methods can address fundamental questions that cannot easily be tackled experimentally [8–10]. Recent years have seen an explosion in the number of modelling studies of enzyme reactions and a transformation in their sophistication and reliability. As well as mechanisms, modelling can address questions of specificity, the effects of mutations or genetic variation and the derivation of SARs [11].

Goals in modelling an enzyme reaction

Taking experimental data, particularly structural information, as a starting point, modelling can tackle questions that are hard to answer experimentally [12]. An essential first step in studying an enzyme-catalyzed reaction is to establish its chemical mechanism. This means determining the identities of catalytic residues, which are usually not obvious, and their roles. Modelling can analyze transition states directly; transition states are central in catalysis and cannot be studied experimentally because they are extremely short lived. Any specific interaction that stabilizes transition states or intermediates should be identified. These interactions might not be apparent from experimental structures and can potentially offer enhanced affinity if they are exploited in designed ligands, as many enzymes show exceptionally high affinities for transition states and intermediates. Calculations can pinpoint catalytically vital functional groups; several examples of key catalytic interactions identified through calculations now exist.

Many enzymes show large conformational changes during their reaction cycles, and the function and relationship of these changes to the chemical steps should be explored [13]. Proteins have complex dynamics, exhibiting a wide range of internal motions, some of which are vital to their function. There have been many suggestions that protein dynamics is important in catalysis, although the simulations indicate that the direct effect of protein dynamics in determining enzyme chemical reaction rates is generally relatively small [8,10,14]. Nevertheless, it is certainly important to consider a representative sample of possible conformations and the effects of conformational variations in enzyme reactions, for example through dynamics or Monte Carlo simulations. Quantum effects, such as nuclear tunnelling, are important in many enzyme reactions involving hydrogen transfer [15,16]. Finally, when the interest lies in understanding why an enzyme is an effective catalyst (i.e. why the enzymatic reaction is faster than its uncatalyzed counterpart) the enzymatic reaction should be compared to an equivalent reaction in solution. Overall, understanding enzyme activity involves many different levels of complexity and presents a range of challenges, for which different theoretical approaches can be best suited, depending on the exact question of interest in a drug design problem.

Methods for modelling enzyme reactions

A central challenge in modelling enzyme reactions is the sheer size of enzymes. Standard 'molecular-mechanics' (MM) methods (e.g. the well-known force fields developed for AMBER, CHARMM and GROMOS) can deal with protein structure and dynamics, but are not applicable to chemical reactions. This is because of their simple functional forms (e.g. harmonic terms for bond stretching) and

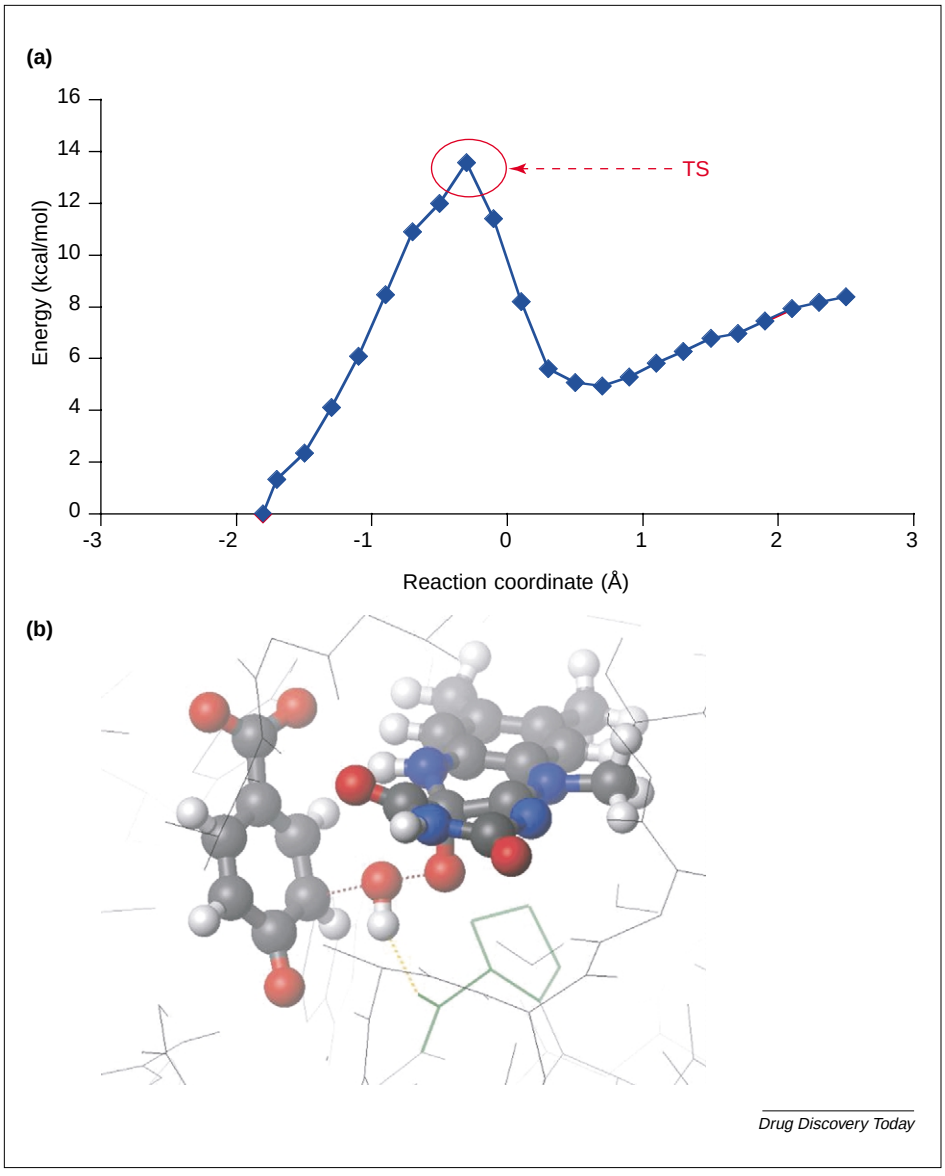

FIGURE 2

Transition-state stabilization in enzyme reactions. Enzymes often show high affinity for binding transition states and reaction intermediates. Modelling can identify the interactions involved, knowledge that can help in designing ligands with increased affinity as pharmaceutical leads. **(a)** Transition-state (TS) stabilization in the enzyme chorismate mutase: combined quantum-mechanics and molecular-mechanics (AM1/CHARMM QM/MM) modelling of the reaction in the enzyme shows that the stabilization provided by the enzyme is maximal for the transition state, indicating their complementarity. The transition state is stabilized by specific electrostatic interactions at the active site [26]. **(b)** Identification of a key catalytic interaction in a flavoprotein monooxygenase: QM/MM modelling of aromatic hydroxylation in *para*-hydroxybenzoate hydroxylase showed the transition state to be stabilized by a hydrogen bond (dotted yellow) with the carbonyl of a conserved proline residue (green) [24].

inability to model changes in electronic polarization. It is possible to develop MM parameters for a specific reaction, but this is highly laborious and these parameters might not be applicable to other cases. Also, the form of the potential function can impose severe limitations, such as neglecting the electronic polarization. A description that takes into account the fundamental quantum mechanics of electronic structure is more general and often easier to apply for reactions. Quantum-chemical methods, based on *ab initio* molecular-orbital or density-functional theory, can currently be used to study reactions in molecular systems containing tens of atoms. Small 'cluster' models of approximately this size can represent key features of an enzyme reaction and can identify possible mechanisms. This is particularly true for metalloenzymes, where all the important chemical steps take place at one metal centre, where the metal holds its ligands in place, limiting the need for artificial restraints to maintain the enzyme structure.

Reliable calculations have increasingly been made possible by the development of methods based on density-functional theory. Widely used functionals (e.g. the B3LYP hybrid functional) give good results for many reactions in a reasonable amount of time. The work of Siegbahn and collaborators exemplifies the biochemical insight that calculations on small models can provide, for example in identifying enzyme mechanisms [17]. Another example, cytochrome P450, is discussed below. Examples of important functional groups are amino-acid side chains involved in catalysis or binding, and the substrate, with their positions taken from a representative X-ray crystal structure of an enzyme complex. It can then be possible to optimize the geometries of complexes representing the reactants, transition state, intermediates and products of the reaction, although a small model might not contain all the important functional groups. Environmental effects, such as solvation and long-range electrostatic interactions, can be included in an approximate way via continuum solvation models, but these cannot fully represent the heterogeneous enzyme environment [10]. An important practical consideration is that it might be difficult to optimize the geometry of the model (for example to locate a transition-state structure) while maintaining the correct orientations of the groups in the protein.

More approximate quantum-chemical methods (e.g. the semiempirical molecular-orbital techniques AM1 and PM3) can model larger systems (hundreds of atoms) but are often inaccurate and in some cases not easy to use (e.g. for many transition metals). Techniques have been developed that allow semiempirical electronic-structure calculations on whole proteins [18–20] and efforts are being made to improve the scaling properties of higher-level quantum-chemical methods, to extend them to larger systems. However, solvated enzyme complexes can

GLOSSARY

Ab initio quantum-mechanical methods are becoming a valuable tool in studying not only atoms and small molecules but also biological molecules. They are quantum-mechanical methods, sometimes described as quantum chemical. Literally, *ab initio* means 'from the beginning' – this signifies that *ab initio* methods aim to work from basic physical principles. These methods calculate the electronic distribution and other properties in a molecule. A wide range of methods, of varying accuracy and computational requirements, have been developed – the best are extremely computationally demanding. Currently, typical *ab initio* methods can be applied to reactions in molecular systems containing tens of atoms. The large computational cost precludes *ab initio* calculations on reactions in whole enzyme–substrate complexes (which typically contain thousands of atoms).

Cluster (or 'supermolecule') models include only a small number of essential groups (e.g. the substrate and key catalytic groups). They are the simplest models of an enzyme-catalyzed reaction and are used because it is not possible to study reactions in larger models by quantum-mechanical calculations.

Density-functional theory (DFT) methods are now among the most popular type of quantum-mechanical (electronic-structure) calculations for molecular systems. The energy is calculated from the electron density. Developments in theories, software and hardware during the 1990s enabled reliable DFT calculations for many molecules and their reactions. Hybrid DFT methods include a more traditional, Hartree–Fock, molecular-orbital, quantum-chemical calculation to calculate part of the energy. DFT methods are available 'off the shelf' in several popular quantum-chemical software packages.

Empirical valence bond (EVB) methods have been used to study many enzyme reactions. Unlike QM/MM calculations, EVB methods do not use a quantum-chemical calculation – the distribution of the electrons is not treated explicitly. Instead, a small number of chemical structures (valence bond structures) is usually used to represent the important species in a reaction or reaction step (e.g. the reactants and products). Each structure is modelled by a specialized molecular-mechanics type of energy function. The total energy of the system during the reaction is found as a 'mixture' of the energy of the different possible structures. EVB calculations can give detailed insight into the energetics of an enzyme reaction, but careful parameterization of the energy functions (e.g. against experimental data) is required for each new system.

Molecular-dynamics simulation methods for proteins and other biological macromolecules model molecular motions, typically using the laws of classical dynamics. Molecular-dynamics simulations can be performed with molecular-mechanics methods or with quantum-mechanics and molecular-mechanics techniques.

Molecular mechanics (MM) is the name given to methods which treat a molecule as 'balls' joined by springs. These methods are empirical, which means they are not based on fundamental physics, instead they contain parameters optimized to reproduce experimental data. Simple functions are used to calculate the energy for stretching bonds and rotating torsion angles, for example. Being relatively simple, molecular-mechanics calculations are fast and can be used for large molecules, such as proteins. However, typical protein molecular-mechanics methods cannot model chemical reactions. The parameters and potential function of a molecular-mechanics method are described as a force field. Several reliable molecular-mechanics force fields have been developed for proteins.

Pharmacogenomics aims to examine variations in genes that might affect drug efficacy and response, and to understand and predict these effects. The aim is to explore how genetic variations can be used to predict whether a patient will benefit from the treatment with a particular drug, might suffer a bad response to a drug (e.g. an adverse drug reaction) or might not benefit from a drug. Variations causing differences in activity and specificity in enzymes that metabolize drugs, such as the cytochromes P450, are of particular interest in this growing area.

Quantum-mechanics (QM) or quantum-chemical methods calculate computationally the distribution of electrons in molecules, using the principles of quantum mechanics, which means treating the electrons as waves, with their distribution described mathematically by wavefunctions. The atomic nuclei (which are much heavier) are usually treated as classical particles, like billiard balls. Quantum-mechanical methods can be used to calculate what happens in a chemical reaction – the distribution of electrons changes in a reaction as bonds are broken and formed. Among the types of quantum-mechanical methods used in chemistry are *ab initio* and semiempirical molecular-orbital calculations and methods based on density-functional theory. The best of these methods can give results very close to experiment, but need large amounts of computer time and resources.

Quantum-mechanics and molecular-mechanics (QM/MM) methods use a quantum-mechanical treatment of the electrons in a small region, for example at the active site of an enzyme. Only a relatively small number of atoms can be included in a quantum-mechanical calculation. In a QM/MM calculation, most of the protein, the solvent and any other groups bound to the protein, are treated by molecular mechanics. To include the effects of the environment on the reaction, the QM and MM regions should interact, which means the QM region should 'feel' the influence of (and be polarized by) the MM atoms. QM/MM methods are sometimes described as hybrid, because they combine QM and MM methods, but this term has other meanings in chemistry, so can be confusing, and in any case QM/MM itself implies a combination of methods.

Semiempirical quantum-mechanical methods apply several approximations to make electronic structure calculations faster (for example, they include only the valence electrons in a molecule). To overcome some of the inaccuracies these approximations cause, semiempirical methods are parameterized to reproduce experimental data for small molecules. They can deal practically with molecular systems containing hundreds of atoms.

Tunnelling is a quantum-mechanical process in which a light particle is able to pass through an energy barrier, instead of going over it. Tunnelling can be important in enzyme reactions involving the transfer of hydrogen atoms or nuclei.

Zero-point corrections. Usually, as mentioned above, atomic nuclei are treated as fixed in a quantum-chemical calculation. In fact, quantum mechanics requires that all bonds vibrate with 'zero-point' energy and this energy should be added to the electronic energy found in a quantum-chemical calculation. When a bond breaks, its zero-point energy is lost. Zero-point energy is large for light atoms, therefore it can be a particularly significant factor in determining the strength of chemical bonds involving hydrogen and the rates at which they break.

contain tens of thousands of atoms, currently even beyond semiempirical methods.

Equally important in modelling a reaction is the need for extensive reaction-pathway calculations and conformational sampling, which are significant challenges for large molecules. The environment of the enzyme (typically aqueous solution, but some enzymes operate in concentrated solutions, in membranes or in protein or nucleic acid complexes) must be considered. Because of the complexity of protein internal motions, many conformational substates can exist and a single structure might not be truly representative [21]. For conformational sampling (e.g. to calculate free energy profiles and potentials of mean force [22]) a simulation method must be capable of calculating trajectories of at least many picoseconds. It can be useful to use MM molecular-dynamics simulations (which can run to relatively long, nanosecond timescales) or to use multiple different crystal structures, to generate multiple models for mechanism calculations and ensure wide sampling of possible enzyme configurations.

Ideally, a method for simulating enzyme reactions should capture the essential details of the chemical reaction, while including the enzyme and solvent. One notable method is the empirical valence bond (EVB) model [23],

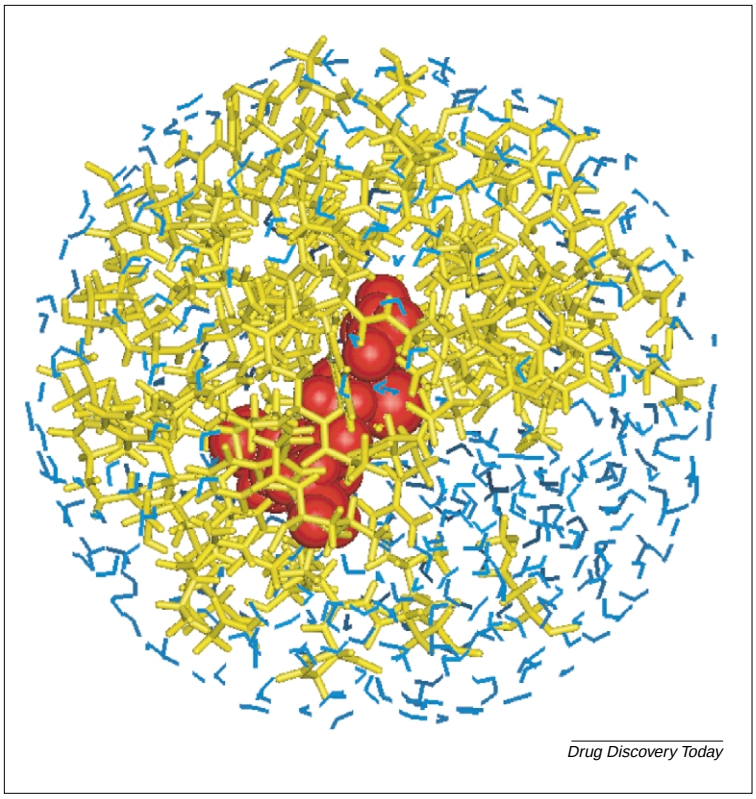

FIGURE 3

QM/MM methods for modelling enzyme reactions. QM/MM methods are an increasingly popular and powerful approach to investigating enzyme mechanisms, specificity and catalysis. The essence of the QM/MM technique is simple: a small region at the active site containing the substrate, catalytic residues and any cofactors (coloured in red) is treated by a quantum-chemical method capable of modelling the making and breaking of bonds. This small region interacts with the protein (in yellow) and solvent environment (in blue), which are treated by a standard empirical molecular-mechanics force field.

in which resonance structures are chosen to represent the reaction, with the energy of each given by a simple empirical force field. The EVB Hamiltonian is calibrated to reproduce experimental data for the reaction in solution, or alternatively *ab initio* results can be used. The free energy of activation for the reaction in solution and with the enzyme can be calculated using free-energy perturbation simulations.

Combined quantum-mechanics and molecular-mechanics methods

Combined quantum-mechanics and molecular-mechanics (QM/MM) methods are increasingly popular and powerful for modelling enzyme reactions. They combine a quantum-chemical description of the groups directly involved in the reaction with a simpler molecular-mechanics treatment of the enzyme and the environment (Figure 3). Different coupling schemes can be used to treat the interaction of the QM and MM regions. Enzymes are polar, so it is probably important to include polarization of the QM atoms by their MM neighbours, as included in most QM/MM enzyme studies. Interest in QM/MM methods has grown rapidly in recent years and it is now clear that they can provide biochemically useful insight into the mechanisms of enzymatic reactions [8–12]. They have demonstrated their value in identifying catalytic functions for active-site residues (such as a conserved proline in flavin-dependent monooxygenases (Figure 2b) [24]), in addressing mechanistic questions and suggesting and testing catalytic principles (such as conformational effects and transition-state stabilization in chorismate mutase (Figure 2a) [25,26,12] – lively current debates on the origin of catalysis in this apparently simple enzyme reaction show how modelling can help to formulate and test mechanisms and hypotheses). Many different QM/MM implementations are available in several widely used program packages. QM/MM calculations can be carried out at different levels of QM electronic structure calculation, such as *ab initio* [24,27,28], semiempirical [29,30], density-functional [31] or approximate density-functional (e.g. the self-consistent charge density-functional tight-binding (SCC-DFTB) [32] method combines computational efficiency with reasonable accuracy for many applications). Transition-state structures can be optimized and molecular-dynamics simulations can be carried out with more approximate, less computer-intensive, QM/MM methods (such as semiempirical QM/MM). Free-energy differences, such as activation free energies, can be calculated, as well as quantum effects, such as tunnelling and zero-point corrections [8,22]. These methods have an important role as they allow more extensive simulations to be performed (e.g. dynamics or Monte Carlo simulations, conformational sampling and reaction pathway calculations). Specifically parameterized semiempirical methods can give improved accuracy for a particular reaction [5,9,22]. High-level QM/MM calculations (e.g. *ab initio* or density-functional level QM) are highly computationally demanding but are required for some systems and can be used to test more approximate approaches.

Building models for simulating enzyme reactions: challenges and pitfalls

Applying QM/MM methods does need care and attention to detail, for example in the choice of the QM system and the QM/MM partitioning [33]. For a small number of enzymes (e.g. chorismate mutase [12,25,26,28]) there is no covalent interaction between the enzyme and the substrate and the separation into QM and MM regions is straightforward; the QM/MM interaction would only include MM terms for van der Waals interactions [34] and QM/MM electrostatic interactions, although in most cases the boundary between the QM and MM regions must separate covalently bonded atoms. Several methods have been developed to allow this QM/MM separation, including 'link' atoms (e.g. a hydrogen atom added to the boundary QM atom) [29,35], generalized hybrid orbitals [36,37], 'pseudobonds' and localized orbitals [30]. These methods have been extensively tested and it has been found that when applied sensibly (e.g. partitioning C–C single bonds

far from the site of chemical change) most can give reasonable descriptions [33]. As important as the choice of the QM/MM partitioning method is the location of the boundary, the treatment of QM/MM bonded interactions (e.g. all MM bonded terms involving at least one MM atom might be retained) and the treatment of QM/MM electrostatic interactions at the boundary (e.g. it is often advisable to remove QM/MM interactions with any covalently bonded MM atoms). Long-range electrostatic effects might be significant and might need to be included (e.g. via continuum solvation models in combination with QM/MM modelling).

At first sight, these factors could make a QM/MM calculation appear intimidating. The application of QM/MM method is not, as yet, as standardized as for purely MM or purely QM calculations. However, extensive work has determined procedures and approaches that have been demonstrated to be reliable and have been shown to make useful predictions. In some cases, it has been possible to validate calculated QM/MM activation energies through correlations with experimental rate constants for enzyme reactions [38]. It is also important to bear in mind the usual complications that can arise in protein modelling (e.g. quality and suitability of the crystal structure, effects of disorder, such as alternative conformations or missing residues, and likely protonation states of ionizable groups). In practice, a high-resolution crystal structure of a relevant enzyme complex (e.g. with a bound substrate or transition-state analogue) is needed for reliable mechanistic modelling [33]. The resolution of the crystal structure is an indication of the precision of the structure but it is certainly not the only factor to consider: the complex must represent the reactive complex and not for example an unreactive conformation, in conditions of pH and solute concentration relevant to the environment in which the enzyme actually functions. Crystal structures represent an average over all the molecules in the crystal and over the time course of the crystallographic experiment; this averaging manifests itself in the alternative conformations sometimes found for some amino acid sidechains. A protein crystal structure should not be viewed as a simple structure of a single molecule. As with any biomolecular simulation, the starting structure should be examined carefully in the early stages for potential difficulties or uncertainties. Expert advice from a structural biologist can help to identify and iron out potential problems.

When embarking on simulation of an enzyme mechanism, it is very important to choose the protonation states of protein groups correctly. The pK_as of acidic and basic amino acid sidechains in enzymes can be significantly different from expected values in solution. Choosing protonation states based on standard pK_a values might give an incorrect model, which in the worst case could lead to the wrong mechanism being modelled. It can be a good idea to use methods of pK_a prediction, particularly for active-site residues.

Applied carefully, QM/MM calculations can give unique insight into enzyme mechanisms and specificity. QM/MM methods are also increasingly important in studies of ligand binding, where they offer several advantages over MM methods, including a better physical description of a ligand (e.g. its polarization) and avoiding the need for time-consuming parameterization. The field is still evolving and it is not yet at the stage where quantitative, exact predictions of, for example, reaction rates can routinely be made. For this reason, it is important to validate predictions from modelling with experimental data: prediction of pK_as of functional groups in proteins provides a useful and accurate example of this type of test. Current developments include the use of QM/MM methods in free-energy perturbation simulations [34], for example to calculate relative binding affinities, and in molecular docking [39]. QM/MM methods will undoubtedly become more important in practical drug design applications.

Examples of recent applications

A key decision in modelling an enzyme reaction is the choice of an appropriate method for the questions of interest, which should be practical in terms of delivering a reliable result in a reasonable time. The capabilities of current methods are best illustrated by some recent applications to enzymes of pharmacological interest.

Cytochrome P450

Recent work on aromatic hydroxylation by cytochrome P450 shows how quantum-chemical calculations on small models can help in developing structure–reactivity relationships. Cytochrome P450 catalyzes a variety of oxidation reactions in many vital processes, such as in carcinogenesis and drug metabolism [40]. A particularly important reaction in drug metabolism [41] is hydroxylation of C–H bonds. This type of reaction can activate prodrugs or influence the bioavailability of pharmaceuticals. For ADME–TOX predictions, the aim is to develop SARs to predict conversions of drugs [1]. Approaches based only on the structures and properties of substrates have proved to be of limited use. More-detailed models are needed to account for the reaction mechanism and specificity of different cytochrome P450 isozymes. P450s have complicated reaction cycles, with several steps, each of which can be rate-limiting, depending on the particular substrate and enzyme. Oxidation is not always the rate-limiting step, but for understanding regioselectivity and predicting ratios of alternative products from a given substrate for example, structure–reactivity relationships for oxidation might be useful.

Bathelt *et al.* [3,4] have studied reactions of simple aromatic compounds with Compound I [a haem oxoiron (IV) porphyrin radical cation, thought to be the active species in oxidation] in a model consisting of the porphyrin without side chains and with the iron coordinated by a methyl mercaptide group (MeS^-, representing cysteinate)

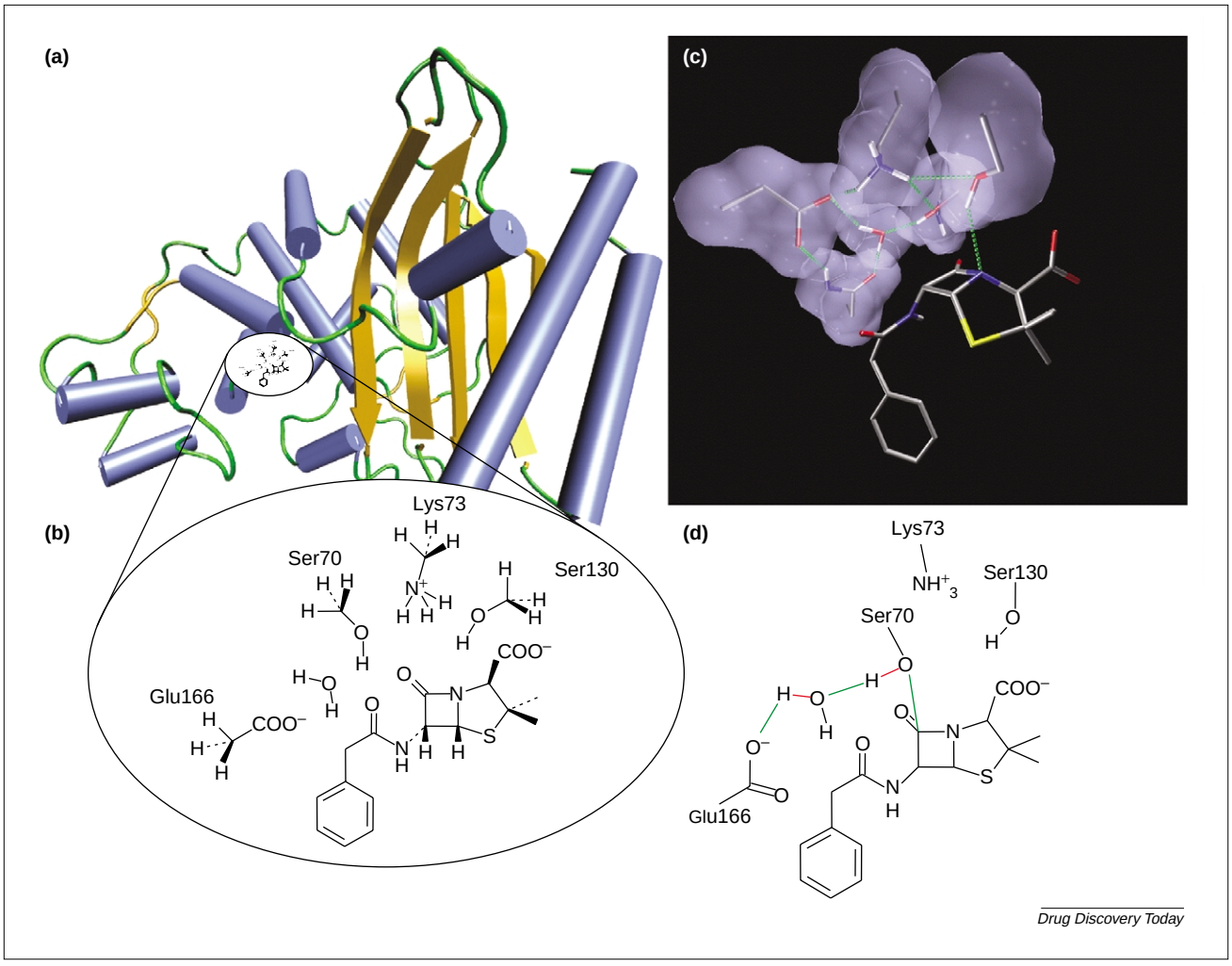

FIGURE 4

Modelling of antibiotic breakdown by a Class A β-lactamase. Understanding the mechanisms of enzyme targets should help in the development of new therapies. The acylation mechanism of benzylpenicillin with a Class A β-lactamase, an important cause of bacterial antibiotic resistance, has been identified by combined quantum-mechanics and molecular-mechanics (QM/MM) modelling [46]. **(a)** The enzyme modelled was the TEM1 β-lactamase from *E. coli*. The simulation system consisted of all residues within 18 Å of the active site. **(b)** Important active site residues were included in the QM region (70 atoms in total). **(c)** Structure of the Michaelis (substrate) complex, optimized at the AM1/CHARMM QM/MM level, showing some important residues. **(d)** The reaction was modelled by defining approximate reaction coordinates in terms of combinations of key interatomic distances (shown in colour). The results identified Glu166 as the base in acylation, acting via a structurally conserved water molecule. Glu166 deprotonates the water molecule, which in turn activates Ser70 as the nucleophile for attack on the β-lactam antibiotic.

(Figure 1). The B3LYP hybrid density-functional theory method used has been shown to predict accurately structures and energetics for bioinorganic systems such as P450 Compound I and other transition-metal complexes. Two different reaction routes were found, differing in the orientation of the benzene ring as it approaches Compound I ('side on' and 'face on', the second with a lower barrier). Both orientations might be important in the reactions of different drugs. The transition state was found to have mixed radical and cationic character, and this insight led to the development of new structure–reactivity relationships (SARs to predict how small molecules will react and be transformed by particular enzymes in metabolism) for a range of substituted benzenes, using a dual-parameter approach combining radical and cationic electronic descriptors.

The reactive properties of the haem group and Compound I are intrinsically similar for all CYP450s. However, substrate specificity and positions of hydroxylation differ widely between different P450 isoenzymes and can be dominated by orientation or affinity effects in the binding site [42] or by the chemical reactivity of different positions of the substrates. Genetic polymorphisms can also be important, for example in drug metabolism [2]. The reactivity of P450s could also be affected by modulation of the electronic properties of Compound I by the protein environment. To examine specificity questions like this, models that take the protein into account are needed. Important steps in this direction have been taken with QM/MM studies of bacterial $P450_{cam}$, which have raised controversial reactivity issues [43,44]. With the recent solution of the crystal structures of human P450s,

QM/MM studies of enzymes directly relevant to drug metabolism are now possible [45].

Mechanisms of antibiotic resistance: class A β lactamase

QM/MM methods have recently been used to model the mechanism of antibiotic breakdown in one of the most important class of bacterial resistance enzymes, Class A β-lactamase [46]. The acylation of the *Escherichia coli* TEM1 lactamase with benzylpenicillin has recently been modelled by QM/MM methods (Figure 4). The QM region (70 atoms) included all active site residues directly involved in the reaction and the substrate. Structures and interactions with the protein were modelled by the AM1/CHARMM22 QM/MM approach, because of the high computational cost of higher level QM/MM modelling, whereas calculation of reaction energies at a better level (B3LYP/6–31+G(d)) corrected for limitations of the semiempirical AM1 method. Potential energy surfaces were calculated by QM/MM energy minimization, using defined reaction coordinates to test possible mechanisms. The results show Glu166 as the general base, deprotonating a structurally conserved water molecule, which in turn activates Ser70 for nucleophilic attack on the antibiotic. Subsequently, the acylenzyme is formed, with Ser130 as the proton donor to the antibiotic thiazolidine ring, with Lys73 acting as a proton shuttle. This mechanism is consistent with experimental kinetic and structural data. For example, the calculated QM/MM energy barrier (9 kcal mol^{-1}) is consistent with the experimental activation energy (12 kcal mol^{-1}). Important interactions were identified, which could help in the development of stable β-lactam antibiotics and in designing new lactamase inhibitors.

Predicting the effects of genetic variation: glutathione-S-transferase

QM/MM molecular-dynamics simulations have been used to model the conjugation of glutathione to phenanthrene 9,10-oxide in a glutathione-S-transferase (GST), pinpointing a determinant of stereospecificity in this epoxide ring opening [5]. A single amino acid difference, not obvious from structural or sequence data, was identified as important in determining the stereospecificity of GST isoenzymes for this reaction. Optimized AM1 parameters were used for the sulphur nucleophile. Umbrella-sampling molecular-dynamics (using the CHARMM program, and CHARMM22 MM protein parameters) simulations were used to calculate free-energy profiles for the reaction, using an approximate reaction coordinate (the difference between the $S_{thiolate}$–$C_{epoxide}$ and $C_{epoxide}$–$O_{epoxide}$ distances). The barriers agree well with experiment. Differences between GST isoenzymes were analyzed by calculating the effects of mutations on barrier heights and reaction energy in the formation of the two possible diastereomeric products, illustrating the potential of QM/MM modelling for understanding and predicting the phenotypic consequences of genetic variations.

Other recent applications

Similar QM/MM methods (AM1/CHARMM and AM1/OPLS-AA) applied to influenza neuraminidase [47] did not support a covalent intermediate but indicate a direct hydroxylation mechanism. However, only a small energy difference was found between these two paths, suggesting that it might be possible to design new covalent inhibitors.

Other interesting recent modelling studies of enzyme reactions include density-functional modelling of the mechanisms of 4-hydroxyphenylpyruvate dioxygenase [48]; naphthalene dioxygenase [49] and class III ribonucleotide reductase [50]; PM3/CHARMM QM/MM modelling of the formation of the Meisenheimer intermediate in 4-chlorobenzoyl-CoA dehalogenase [51]; proline isomerization in cylcophilin and mutants with SCC-DFTB/CHARMM QM/MM [52]; QM/MM calculations of kinetic isotope effects in chorismate mutase [53] and catechol O-methyltransferase [54]; QM/MM dynamics and density-functional theory studies of a metallo β-lactamase [55]; QM/MM study of the contribution of the protein backbone in the mechanism of 4-oxalocrotonate tautomerase [56]; density-functional QM/MM calculation of the free-energy profile in chorismate mutase by multiple steered molecular-dynamics simulations [57]; QM/MM Monte Carlo free-energy perturbation simulations of macrophomate synthase [58]; QM/MM modelling of inhibition mechanisms of neutrophil elastase [59]; demonstration of substrate autocatalysis in uracil DNA-glysosylase [60]; and simulation of redox-coupled proton transfer in cytochrome *c* oxidase [61]. A recent study used a validated homology model to examine the substrate binding mode and reaction mechanism of a malaria protease with a novel active site [62]. This histo-aspartic protease (HAP) from the malaria parasite *Plasmodium falciparum* is a target for drug design but its 3D structure is not known. Bjelic and Åqvist [62] used a combination of homology modelling, automated docking, and molecular-dynamics and reaction free-energy profile simulations to predict the structure of the enzyme and the conformation of the bound substrate, and its catalytic mechanism. The reaction mechanism was found to involve directly only a catalytic aspartate and was stabilized by a histidine residue. The calculated reaction rate agreed well with experimental kinetic data for a hexapeptide substrate derived from human haemoglobin.

Conclusions

The availability of new powerful methods makes modelling and simulation excellent ways of investigating enzyme activity, specificity and catalysis. Given the complexity of enzymes, this should always be done with care and detailed testing. These techniques can then provide uniquely detailed analysis and useful practical insight into key enzymological problems in drug discovery, and their contribution will increase in the future.

Acknowledgements

Adrian J. Mulholland thanks Christine Bathelt, Kara Ranaghan, Johannes Hermann, Lars Ridder, Jeremy Harvey and Anna Bowman for useful discussions and assistance with preparation of the Figures. Adrian J. Mulholland also thanks BBSRC, EPSRC, Vernalis and the IBM High Performance Computing Life Sciences Outreach Programme for support.

References

1 Ridder, L. and Mulholland, A.J. (2003) Modeling biotransformation reactions by combined quantum mechanical/molecular mechanical approaches: From structure to activity. *Curr. Top. Med. Chem.* 3, 1241–1256
2 Pirmohamed, M. and Park, B.K. (2003) Cytochrome P450 enzyme polymorphisms and adverse drug reactions. *Toxicology* 192, 23–32
3 Bathelt, C.M. *et al.* (2003) Aromatic hydroxylation by cytochrome P450: model calculations of mechanism and substituent effects. *J. Am. Chem. Soc.* 125, 15004–15005
4 Bathelt, C.M. *et al.* (2004) Mechanism and structure–reactivity relationships for aromatic hydroxylation by cytochrome P450. *Org. Biomol. Chem.* 2, 2998–3005
5 Ridder, L. *et al.* (2002) Quantum mechanical/molecular mechanical free energy simulations of the glutathione *S*-transferase (M1-1) reaction with phenanthrene 9,10-oxide 124. *J. Am. Chem. Soc.* 124, 9926–9936
6 Cunningham, M.A. and Bash, P.A. (1997) Computational enzymology. *Biochimie* 79, 687–689
7 Bruice, T.C. and Kahn, K. (2000) Computational enzymology. *Curr. Opin. Chem. Biol.* 4, 540–544
8 Garcia-Viloca, M. *et al.* (2004) How enzymes work: analysis by modern rate theory and computer simulations. *Science* 303, 186–195
9 Field, M.J. (2002) Simulating enzyme reactions: challenges and perspectives. *J. Comput. Chem.* 23, 48–58
10 Shurki, A. and Warshel, A. (2003) Structure/function correlations of proteins using MM, QM/MM, and related approaches: methods, concepts, pitfalls, and current progress. *Adv. Protein Chem.* 66, 249–313
11 Perruccio, F. *et al.* (2003) Quantum-mechanical/molecular-mechanical methods in medicinal chemistry. In *Quantum Medicinal Chemistry* (Carloni, P. and Alber, F. eds), pp. 177–198, Wiley-VCH
12 Marti, S. *et al.* (2004) Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.* 33, 98–107
13 Karplus, M. *et al.* (2005) Protein structural transitions and their functional role. *Philos. Transact. A Math. Phys. Eng. Sci.* 363, 331–355
14 Olsson, M.H.M. and Warshel, A. (2004) Solute solvent dynamics and energetics in enzyme catalysis: The S_N2 reaction of dehalogenase as a general benchmark. *J. Am. Chem. Soc.* 126, 15167–15179
15 Kohen, A. *et al.* (1999) Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase. *Nature* 399, 496–499
16 Masgrau, L. *et al.* (2004) Hydrogen tunneling in quinoproteins. *Arch. Biochem. Biophys.* 428, 41–51
17 Himo, F. and Siegbahn, P.E.M. (2003) Quantum chemical studies of radical-containing enzymes. *Chem. Rev.* 103, 2421–2456
18 Van der Vaart, A. *et al.* (2000) Linear scaling molecular orbital calculations of biological systems using the semiempirical divide and conquer method. *J. Comp. Chem.* 21, 1494–1504
19 Khandogin, J. and York, D.M. (2004) Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods. *Proteins* 56, 724–737
20 Khandogin, J. *et al.* (2003) Insights into the regioselectivity and RNA-binding affinity of HIV-1 nucleocapsid protein from linear-scaling quantum methods. *J. Mol. Biol.* 330, 993–1004
21 Zhang, Y.K. *et al.* (2003) Influence of structural fluctuation on enzyme reaction energy barriers in combined quantum mechanical/molecular mechanical studies. *J. Phys. Chem. B* 107, 4459–4463
22 Gao, J. and Truhlar, D.G. (2002) Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* 53, 467–505
23 Warshel, A. (2003) Computer simulations of enzyme catalysis: methods, progress, and insights. *Annu. Rev. Biophys. Biomol. Struct.* 32, 425–443
24 Ridder, L. *et al.* (2003) Ab initio QM/MM modeling of the hydroxylation step in p-hydroxybenzoate hydroxylase. *J. Phys. Chem. B* 107, 2118–2126
25 Ranaghan, K.E. and Mulholland, A.J. (2004) Conformational effects in enzyme catalysis: QM/MM free energy calculation of the 'NAC' contribution in chorismate mutase. *Chem. Commun. (Camb)* 1238–1239
26 Ranaghan, K.E. *et al.* (2004) Transition state stabilization and substrate strain in enzyme catalysis: *ab initio* QM/MM modelling of the chorismate mutase reaction. *Org. Biomol. Chem.* 2, 968–980
27 Mulholland, A.J. *et al.* (2000) Ab initio QM/MM study of the citrate synthase mechanism. A low-barrier hydrogen bond is not involved. *J. Am. Chem. Soc.* 122, 534–535
28 Woodcock, H.L. *et al.* (2003) Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theor. Chem. Acc.* 109, 140–148
29 Field, M.J. *et al.* (1990) Combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comp. Chem.* 11, 700–733
30 Warshel, A. and Levitt, M. (1976) Theoretical studies of enzymic reactions – dielectric, electrostatic and steric stabilization of carbonium ion in reaction of lysozyme. *J. Mol. Biol.* 103, 227–249
31 Lyne, P.D. *et al.* (1999) A hybrid QM-MM potential employing Hartree-Fock or density functional methods in the quantum region. *J. Phys. Chem. A* 103, 3462–3471
32 Cui, Q. *et al.* (2001) A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B* 105, 569–585
33 Mulholland, A.J. (2001) The QM/MM Approach to Enzymatic Reactions. In *Theoretical Biochemistry* (Erikkson, L.A. ed.), pp. 597–653, Elsevier
34 Riccardi, D. *et al.* (2004) Importance of van der Waals interactions in QM/MM Simulations. *J. Phys. Chem. B* 108, 6467–6478
35 Das, D. *et al.* (2002) Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *J. Chem. Phys.* 117, 10534–10547
36 Pu, J.Z. *et al.* (2004) Combining self-consistent-charge density-functional tight-binding (SCC-DFTB) with molecular mechanics by the generalized hybrid orbital (GHO) method. *J. Phys. Chem. A* 108, 5454–5463
37 Pu, J.Z. *et al.* (2004) Generalized hybrid orbital (GHO) method for combining ab initio Hartree-Fock wave functions with molecular mechanics. *J. Phys. Chem. A* 108, 632–650
38 Ridder, L. *et al.* (2000) A quantum mechanical/molecular mechanical study of the hydroxylation of phenol and hydroxylated derivatives by phenol hydroxylase. *J. Am. Chem. Soc.* 122, 8728–8738
39 Raha, K. and Merz, K.M. (2004) A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding. *J. Am. Chem. Soc.* 126, 1020–1021
40 Meunier, B. *et al.* (2004) Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* 104, 3947–3980
41 Guengerich, F.P. (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.* 14, 611–650
42 de Groot, M.J. *et al.* (2004) In silico methods for predicting ligand binding determinants of cytochromes P450. *Curr. Top. Med. Chem.* 4, 1803–1824
43 Schoneboom, J.C. *et al.* (2004) Quantum mechanical/molecular mechanical investigation of the mechanism of C-H hydroxylation of camphor by cytochrome $P450_{cam}$: theory supports a two-state rebound mechanism. *J. Am. Chem. Soc.* 126, 4017–4034
44 Guallar, V. *et al.* (2003) Peripheral heme substituents control the hydrogen-atom abstraction chemistry in cytochromes P450. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6998–7002
45 Bathelt, C. *et al.* (2005) Electronic structure of Compound I in human isoforms of cytochrome P450 from QM/MM modelling. *J. Am. Chem. Soc.* (in press)
46 Hermann, J.C. *et al.* (2005) Mechanisms of antibiotic resistance: QM/MM modeling of the acylation reaction of a class A beta-lactamase with benzylpenicillin. *J. Am. Chem. Soc.* 127, 4454–4465
47 Thomas, A. *et al.* (1999) Is there a covalent intermediate in the viral neuraminidase reaction? A hybrid potential free-energy study. *J. Am. Chem. Soc.* 121, 9693–9702
48 Borowski, T. *et al.* (2004) 4-hydroxyphenylpyruvate dioxygenase: a hybrid density functional study of the catalytic reaction mechanism. *Biochemistry* 43, 12331–12342
49 Bassan, A. *et al.* (2004) A theoretical study of the cis-dihydroxylation mechanism in naphthalene 1,2-dioxygenase. *J. Biol. Inorg. Chem.* 9, 439–452
50 Cho, K.B. *et al.* (2004) Density functional calculations on class III ribonucleotide

reductase: substrate reaction mechanism with two formates. *J. Phys. Chem. B* 108, 2056–2065
51 Xu, D. *et al.* (2004) QM/MM studies of the enzyme-catalyzed dechlorination of 4-chlorobenzoyl-CoA provide insight into reaction energetics. *J. Am. Chem. Soc.* 126, 13649–13658
52 Li, G. and Cui, Q. (2003) What is so special about Arg 55 in the catalysis of cyclophilin A? Insights from hybrid QM/MM simulations. *J. Am. Chem. Soc.* 125, 15028–15038
53 Marti, S. *et al.* (2005) Computing kinetic isotope effects for chorismate mutase with high accuracy. A new DFT/MM strategy. *J. Phys. Chem. B* 109, 3707–3710
54 Ruggiero, G.D. *et al.* (2004) QM/MM determination of kinetic isotope effects for COMT-catalyzed methyl transfer does not support compression hypothesis. *J. Am. Chem. Soc.* 126, 8634–8635
55 Park, H. *et al.* (2005) Hybrid QM/MM and DIFT investigations of the catalytic mechanism and inhibition of the dinuclear zinc metallo-beta-lactamase CcrA from *Bacteroides fragilis*. *J. Am. Chem. Soc.* 127, 4232–4241
56 Cisneros, G.A. *et al.* (2004) The protein backbone makes important contributions to 4-oxalocrotonate tautomerase enzyme catalysis: understanding from theory and experiment. *Biochemistry* 43, 6885–6892
57 Crespo, A. *et al.* (2005) Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase. *J. Am. Chem. Soc.* 127, 6940–6941
58 Guimaraes, C.R.W. *et al.* (2005) Macrophomate synthase: QM/MM simulations address the Diels–Alder versus Michael–Aldol reaction mechanism. *J. Am. Chem. Soc.* 127, 3577–3588
59 Gleeson, M.P. *et al.* (2004) Theoretical analysis of peptidyl alpha-ketoheterocyclic inhibitors of human neutrophil elastase: Insight into the mechanism of inhibition and the application of QM/MM calculations in structure-based drug design. *Org. Biomol. Chem.* 2, 2275–2280
60 Dinner, A.R. *et al.* (2001) Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* 413, 752–755
61 Olsson, M.H.M. *et al.* (2005) Simulating redox coupled proton transfer in cytochrome *c* oxidase: Looking for the proton bottleneck. *FEBS Lett.* 579, 2026–2034
62 Bjelic, S. and Åqvist, J. (2004) Prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site. *Biochemistry* 43, 14521–14528