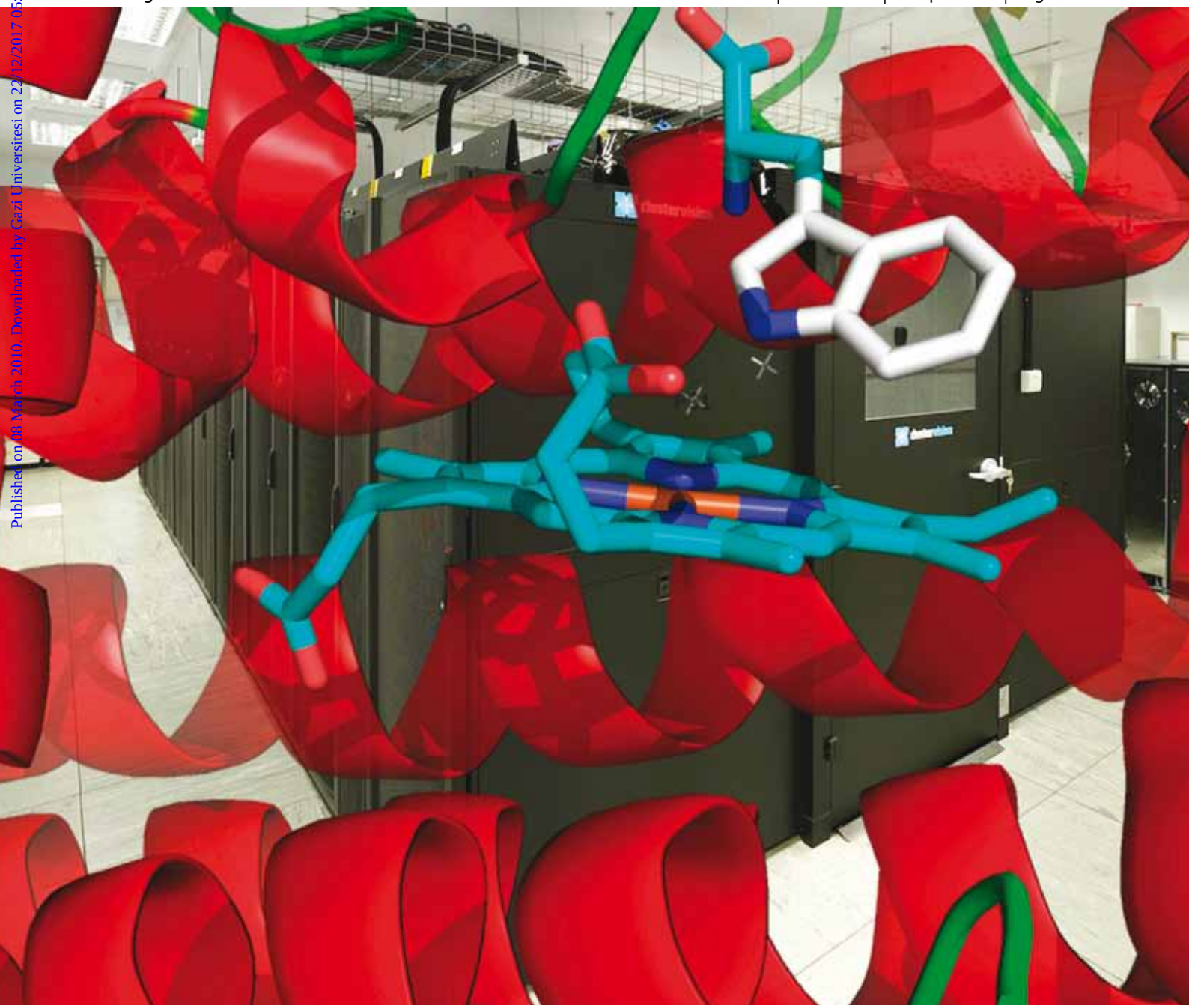
# ChemComm

## Chemical Communications

www.rsc.org/chemcomm

Volume 46 | Number 14 | 14 April 2010 | Pages 2333–2512





ISSN 1359-7345

RSC Publishing

**FEATURE ARTICLE**
Adrian J. Mulholland *et al.*
Computational enzymology

# Computational enzymology

**Richard Lonsdale, Kara E. Ranaghan and Adrian J. Mulholland***

***Received (in Cambridge, UK) 8th December 2009, Accepted 22nd February 2010***
***First published as an Advance Article on the web 8th March 2010***
**DOI: 10.1039/b925647d**

Molecular simulations and modelling are changing the science of enzymology. Calculations can provide detailed, atomic-level insight into the fundamental mechanisms of biological catalysts. Computational enzymology is a rapidly developing area, and is testing theories of catalysis, challenging 'textbook' mechanisms, and identifying novel catalytic mechanisms. Increasingly, modelling is contributing directly to experimental studies of enzyme-catalysed reactions. Potential practical applications include interpretation of experimental data, catalyst design and drug development.



## 1. Introduction

Ever since they were first recognised, the outstanding catalytic properties of enzymes have inspired awe and envy in chemists. Understanding how enzymes 'work'—how they achieve great rate accelerations, efficiency and specificity, under typically mild conditions—is a fundamentally important question in biology and will also have practical benefits. Understanding enzyme catalytic mechanisms at the atomic level could contribute to a range of technological applications such as the design of inhibitors as pharmaceutical leads, prediction of drug metabolism, and design of catalysts for specific transformations. Determination of enzyme mechanisms has proved difficult in many cases, as differentiation between alternative possibilities is often challenging. There is evidence to suggest that several 'textbook' mechanisms are incorrect in important details (hen egg-white lysozyme, for example[1,2]). Enzymology has been notable for many long and heated debates about mechanism and catalytic principles, which have proved very difficult to resolve through experiment alone. There have been many controversial proposals seeking to explain catalysis in some or all enzymes; arguments have centred on the possible role of 'low-barrier' hydrogen bonds,[3–6] so-called 'near-attack conformations',[7–9] enzyme dynamics,[10–12] quantum tunnelling,[12–16] reorganization energy[17] and entropic effects.[18] The applicability of transition state theory to enzyme reactions has also been questioned. The complexity of enzymes, and the difficulty of studying reactions in them, make these arguments difficult to resolve through experiments alone.

Modelling and simulation have the potential to give detailed, atomic-level analysis of enzyme mechanisms and catalysis, and of questions such as the effects of mutations and of how chemical and conformational changes are coupled in enzymes. A range of computational methods is now available for the modelling of enzymes and their reactions, complementing experimental approaches. Techniques for modelling enzyme mechanisms have advanced significantly, and are now making an increasingly important practical contribution to enzymology, and biochemistry more widely.[19] A new field of computational enzymology has emerged and is maturing rapidly.[17,20–28] Molecular modelling and simulation[29] will be—and are proving to be—central to resolving these debates, both in the interpretation of experimental data, and in providing atomic-level analysis of reactions in enzymes. Molecular simulations can probe

*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol, UK BS8 1TS. E-mail: Adrian.Mulholland@bris.ac.uk; Fax: +44 (0) 117 925 1295; Tel: +44 (0) 117 928 9097*

**Richard Lonsdale**

*Richard Lonsdale received his MChem in Chemistry from the University of Oxford in 2005. He obtained his PhD in Chemistry from the University of Bristol in 2009, under the supervision of Adrian Mulholland. He is currently a post-doctoral research assistant in the same group, modelling the mechanisms of enzymes involved in drug metabolism.*

**Kara E. Ranaghan**

*Kara Ranaghan obtained her MSci in Chemistry with study in continental Europe from the University of Bristol in 2002 and went on to obtain a PhD in Chemistry in 2006 (also from the University of Bristol). She is currently a post-doctoral research assistant in the Mulholland group, with an interest in modelling enzyme reaction mechanisms, particularly where quantum mechanical tunnelling may play a role.*

enzyme mechanisms, and the origins of catalysis, at a level of detail that cannot—at least for the foreseeable future—be achieved experimentally. Modelling can identify probable chemical mechanisms, analyse the effects of mutations and genetic variations, pinpoint the causes of specificity, and derive structure–activity relationships.

Computational enzymology began in the early to mid-1970s.[30,31] The pioneering studies of Warshel, continuing to the present day, are particularly notable. By the early 1990s,[32–34] the number of computational mechanistic studies of enzymes was relatively small, but recent years have seen an explosion in the number of computational studies of enzymic reaction mechanisms.[23,35–37] The number of published studies is now so large that it is not possible to cover the majority (even of recent studies) in a short review: the aim here is to highlight a small number of illustrative applications. There has also been a transformation in the accuracy of the computational methods. It is now possible to achieve an unprecedented level of accuracy in calculations on enzyme-catalysed reactions with combined quantum mechanics/molecular mechanics (QM/MM, or QM-MM) methods.[38] High-level quantum chemical methods can now be used to study enzymic reactions allowing calculations of energy barriers for enzyme-catalysed reactions, in the best cases, of near 'chemical accuracy' (1 kcal $mol^{-1}$).[39] Quantitative predictions at this level in first principles calculations were only previously possible for very small molecules. Carefully parameterized empirical molecular simulation approaches also give excellent agreement with experiments for enzymic reactions.[40] In the best cases, calculations can give activation energies that agree very well with experiments.[17,38] Such good agreement indicates that transition state theory provides a good general framework for understanding the rates of enzyme-catalysed reactions.[17,21,38]

## 2. Modelling enzyme-catalysed reactions: how and why

An enzyme structure from X-ray crystallography is the usual starting point for modelling an enzyme-catalysed reaction. In some cases a model constructed based on homology to other structures may be sufficiently reliable,[41] though such models should be treated with much more caution. The first step in studying an enzyme-catalysed reaction is then to establish its chemical mechanism. This is far from trivial: as noted above, many 'textbook' mechanisms are probably wrong. An initial goal is to determine the functions of catalytic residues, which are often not obvious. Even the identities of the important groups may not be certain. Any specific interactions that stabilize transition states or reactive intermediates should also be identified and analysed.

Calculations offer several unique advantages: for example, they can analyse transition states directly. Transition states are central to understanding chemical reactivity and catalysis, but experiments cannot directly study them in enzymes because of their extremely short lifetimes, and because of the large size and complexity of enzymes. Calculations can also identify functional groups and interactions involved in catalysis. Several examples have been published of key catalytic interactions that have been identified by modelling (for example, a conserved proline residue that specifically stabilizes the transition state for aromatic hydroxylation in the flavin dependent monooxygenases *para*-hydroxybenzoate hydroxylase[41] and phenol hydroxylase[42,43]). Functionally important interactions of this type may well not be apparent from experimental structures: for example, they may not exist in structures that can be isolated and crystallized. In addition to providing detailed understanding of the reaction in the enzyme, identifying interactions of this type may assist ligand design: such interactions potentially offer enhanced affinity if they can be exploited in designed ligands (*e.g.* pharmaceutical lead compounds), because many enzymes show exceptionally high apparent binding affinities for transition states and reaction intermediates.

Protein dynamics are complex and fascinating. As molecular dynamics simulations and experiments (*e.g.* NMR) have shown, proteins undergo a wide range of internal motions, some of which are central to their biological function.[44] Many enzymes show large conformational changes during their reaction cycles,[45] and the functions and relationship of these changes to the chemical steps in the reaction (to which they may be intimately related) are questions of considerable current interest. It has been suggested that protein dynamics may contribute to enzyme catalysis, but simulations indicate that the direct effect of protein dynamics in determining the rates of chemical reactions in enzymes is generally relatively small.[46] On the other hand, protein conformational changes (*e.g.* involved with substrate binding or product release) may be the rate-limiting factor for the overall reaction in many enzymes.[47] It is certainly important to consider the possible effects of protein conformational fluctuations and variations when modelling enzyme reactions, *i.e.* to consider a representative sample of possible conformations, *e.g.* through molecular dynamics or Monte Carlo simulations. Quantum effects such as nuclear tunnelling are important in many enzyme reactions involving hydrogen transfer[24,48,49] and their role should be considered.

It is worth emphasizing that, to understand why a given enzyme is an effective *catalyst*, *i.e.* to understand why the reaction in the enzyme reaction proceeds more quickly than the uncatalysed reaction, the enzymic and an equivalent

**Adrian J. Mulholland**

*Adrian Mulholland received his BSc in Chemistry from the University of Bristol in 1990. After a short period at ICI Pharmaceuticals, he obtained his DPhil from Oxford University in 1995. He was a Wellcome Trust Fellow at Harvard, then Bristol, and an EPSRC Advanced Research Fellow. He is currently a Professor of Chemistry and EPSRC Leadership Fellow, with interests including biomolecular simulation, protein dynamics and enzyme catalysis.*

('reference') solution phase reaction should be compared (though it may not necessarily be obvious for all enzymes what the appropriate reference should be).[22] In practical terms, though, often the interest is more in being able to predict the effects of a mutation on activity, or on the specificity of an enzyme for alternative substrates. Overall, understanding an enzyme mechanism, specificity and catalysis involves many different levels of complexity. This presents a variety of different challenges, and different types of modelling or simulation methods are needed to investigate different types of question. An outline of various methods used in computational enzymology is given below.

## 3. Methods for modelling enzyme-catalysed reactions

Enzymes are large molecules, which means that modelling the reactions that they catalyze is complex and challenging. This can be complicated further by the need to include part of a particular enzyme's surrounding environment, such as the surrounding solvent, cofactors, other proteins, a lipid membrane, or DNA. There are many practical considerations in simulating such complex systems, such as interpretation of crystal structures, choice of protonation states for ionizable amino acids *etc*.[29] Our focus here is to discuss the methods used.

### 3.1 Empirical valence bond methods

To examine some important questions relating to enzyme action (*e.g.* to analyse the causes of catalysis, *i.e.* why an enzymic reaction proceeds faster than the equivalent, uncatalysed reaction in solution), it is necessary to use a method that not only captures the essential details of the chemical reaction, but also includes the effects of the enzyme and solvent environment. One notable method in this area is the empirical valence bond (EVB) model.[22] The EVB method has been instrumental in the development of the field of computational enzymology. In the empirical valence bond approach, resonance structures (for example ionic and covalent resonance forms) are chosen to represent the reaction. The energy of each resonance form is given by a simple empirical force field (with realistic treatment of stretching important bonds, for example). The potential energy is given by solving the related secular equation. The EVB Hamiltonian is calibrated to reproduce experimental data for a known and relevant solution reaction, or alternatively *ab initio* results can be used.[50] It is assumed that the off-diagonal elements of the EVB Hamiltonian do not differ significantly between the enzyme and solution.[51] This approximation has been tested recently using frozen density functional theory and the constrained density functional theory models to generate convenient diabatic states for QM/MM treatments, and found to be a reasonable assumption for the test case of $S_N2$ reactions.[52] The surrounding protein and solution are modelled by an empirical force field, with appropriate treatment of long-range electrostatics. The free energy of activation for the reaction in solution, and in the enzyme, can be calculated using free energy perturbation simulations.[53]

One of the main advantages of the EVB method is that the free energy surfaces can be calibrated by comparison with experimental data for reference reactions in solution. However, as in any valence bond representation, it is essential that the valence bond forms should represent all the resonance forms that are important in the reaction. An appealing feature of the EVB method is that it makes it straightforward to use a non-geometrical reaction coordinate in modelling a reaction, which may be significantly more accurate for some condensed phase reactions. Energy can be used as a reaction coordinate *e.g.* the energy difference between valence bond structures. A mapping procedure is followed which moves gradually from the reactant to the product. In this mapping, the change in both the solute structure and charge is taken into account. This method locates the correct transition state in the combined solute–solvent reaction coordinate. This allows the evaluation of nonequilibrium solvation effects, for example.[22] The simplicity of the potential function allows extensive sampling in molecular dynamics simulations. Other strengths of the method have been discussed elsewhere.[54] The EVB method is a powerful and useful approach, which has now become a widely adopted tool for studying reactions in condensed phases. Among many investigations with EVB methods are studies of alternative nucleotide insertion mechanisms for T7 DNA polymerase;[55] human aldose reductase;[56] electrostatic effects in catalysis in ketosteroid isomerase;[57] comparison of catalysis in different aspartic proteases;[58] ribosomal peptide bond formation;[59,60] hydride transfer[61] and the temperature dependence of kinetic isotope effects in dihydrofolate reductase;[62] and radical reaction in vitamin $B_{12}$ enzymes, leading to a suggestion of a new paradigm of electrostatic catalysis for such radical reactions.[63] The EVB method could have a useful application in the late stages of computer-aided enzyme design. In a recent study, Warshel *et al.* demonstrated the EVB method to be an accurate and reasonably fast method for calculating transition state free energies in different chorismate mutase enzymes and their mutants.[64] This method could potentially be used to rank the catalytic efficiency of different proposed enzymes that are required to carry out a particular catalytic function.

### 3.2 Quantum chemical calculations on small (active site) models

Quantum chemical calculations (*i.e.* methods that calculate molecular electronic structure, for example *ab initio* molecular orbital or density-functional theory calculations) can provide excellent results for reactions of small molecules. The major problem in applying these methods to model enzyme reactions is that such calculations require very large computational resources (which increase significantly with system size), placing severe practical limitations on the size of the system that can be treated. In most enzymes, however, the chemical changes occurring in the reaction are confined to a relatively small region, the active site of the enzyme. One approach to the study of enzyme-catalysed reactions is therefore to study just the relatively small active site region.[26,28] Studying small models in this way allows the use of powerful (and potentially highly accurate) quantum chemical methods. The best

'*ab initio*' methods, which include the effects of electron correlation, allow calculations of rate constants for reactions involving very few atoms (in the gas phase) with error bars comparable to experiments on these systems.[38] These are highly demanding calculations, however. More approximate methods, (such as the semiempirical molecular orbital techniques AM1[65] and PM3[66]), are relatively computationally cheap, and can model larger systems (containing of the order of hundreds of atoms). Techniques (*e.g.* 'linear-scaling' methods) have been developed that allow semiempirical electronic structure calculations on whole proteins.[67–69] However, semiempirical methods are well known to be inaccurate for many applications (*e.g.* the AM1 and PM3 methods frequently give errors of 10 kcal $mol^{-1}$ or more for calculated barriers and reaction energies[43,70]). They are also unreliable for some structural properties, *e.g.* for peptide bonds and hydrogen bonds.[71] Density functional theory (DFT) methods (*e.g.* the widely-used B3LYP functional) are generally considerably more accurate than semiempirical methods, and permit calculations on relatively large systems (*e.g.* active site models of the order of 100 atoms), larger than is feasible with correlated *ab initio* calculations. DFT methods have opened new classes of enzymes (particularly metalloenzymes such as cytochrome P450 enzymes[20,23,27]) to computational investigation.[26,28] Most DFT methods, however, lack important physical interactions, such as dispersion, which are important in the binding of ligands to proteins. Efforts have been made to compensate for this, for example, by the inclusion of empirical dispersion corrections to existing functionals (*e.g.* DFT-D[72]). DFT methods often give barrier heights that are too low by several kcal $mol^{-1}$,[38] and, more fundamentally, it can be difficult to assess the accuracy of results, because DFT does not offer a route to their systematic improvement or testing.

Small 'cluster' active site models of up to around 100 atoms (sometimes referred to as the 'supermolecule' approach), can represent important features of an enzyme reaction, and identify probable mechanisms. The active site model should contain molecules representing the substrate(s) (and cofactors in cases where they are involved), and protein residues involved in the chemical reaction or binding substrate. The interactions that bind substrates into the active site of an enzyme are typically weak interactions, such as hydrogen bonds, electrostatic and van der Waals interactions. Important functional groups (for example catalytic amino acid side chains) are represented by small molecules. For example, imidazole may be used to represent histidine, acetate to represent an aspartate side chain, and so on. The initial positions of these groups would usually be coordinates extracted from a representative X-ray crystal structure of an enzyme complex, or perhaps from a molecular dynamics simulation of such a (*e.g.* enzyme-substrate or enzyme-intermediate) complex.

Calculations on active site models can examine interactions between groups at the active site, and can provide useful models of transition states and intermediates (see below). They can also be useful to test the accuracy of different levels of calculations for a reaction. This approach has shown itself to be particularly useful for studying the reaction mechanisms of metalloenzymes, for which reliable, semi-quantitatively accurate calculations have been made feasible by the development of methods based on density functional theory. In many metalloenzymes, all the important chemical steps may take place at one metal centre (or a small number of metal ions bound at one site), and the metal may also hold its ligands in place. This gives the technical advantage of limiting the requirement for applying restraints to maintain the correct active site structure.

The work of Siegbahn and collaborators[28,73,74] provides excellent examples of the mechanistic insight that calculations on small clusters can give. For example, such calculations may be able to discriminate between alternative mechanisms: a mechanism can be ruled out if the calculated barriers for it are significantly higher than the experimentally derived activation energy, based on the likely accuracy of the computational method. The energy difference between alternative mechanisms can be very large, larger than the likely effects of the environment (which are either not included, or included only in an approximate way (*e.g.* *via* a continuum solvent model), in calculations on an isolated cluster).

To calculate the energy barrier to reaction in a cluster model, it is necessary to locate the structures (optimize the geometries) of the reactant, transition state, intermediates and product complexes of the reaction. Technically, it can be difficult to optimize the geometry of the model (for example to find a transition state structure), while at the same time maintaining the correct orientations of the groups in the protein. Also, small models may lack important functional groups, and careful thought should be given as to which groups to include, striking a balance between computational feasibility and the desire for a larger, more extensive model. Perhaps counter-intuitively, a larger model is not always a better model: a larger model will be subject to greater conformational complexity (conformational changes distant from the reaction centre may artificially affect relative energies along the reaction path), and charged groups could have unrealistically large effects on reaction energies. It is possible to include the effects of the protein and solvent environment approximately using continuum solvation models, but these cannot fully represent the heterogeneous electrostatic environment in an enzyme.[75]

Enzyme-substrate complexes typically contain thousands of atoms, perhaps tens of thousands, particularly when modelled using an explicit representation of surrounding solvent. This puts them currently beyond even semiempirical quantum chemical methods for modelling reactions. An equally important consideration in reaction modelling is that important structures (such as transition state structures) and preferably entire reaction pathways should be optimized or simulated. The environment of the enzyme (typically aqueous solution, but some enzymes operate in concentrated solutions, in membranes or in protein or nucleic acid complexes) should also be considered. Proteins have many conformational substrates, and a single structure may not be truly representative.[76] Extensive conformational sampling may be needed to generate a representative ensemble of structures.[38,77] These are all significant challenges for large structures. For conformational sampling (*e.g.* to calculate free energy profiles,

*i.e.* potentials of mean force[78]), a useful simulation method must be capable of calculating trajectories of many picoseconds at least (or an equivalently large number of configurations in a Monte Carlo simulation[79]). Molecular dynamics simulations (*e.g.* with MM or low-level QM/MM methods) can be used to generate several structural models for mechanism calculations, to ensure wide sampling of possible enzyme configurations.[38,76,77] If multiple different crystal structures of the same enzyme are available, these may be suitable as different starting models, and similarly help to examine the effects of structural variation on the reaction.[80,81]

'Hybrid' methods that combine quantum chemical methods with molecular mechanics permit extensive calculations on larger models of enzymes than is possible with purely quantum chemical techniques. These QM/MM methods are very important in computational enzymology, and are discussed in detail below.

### 3.3 Combined quantum mechanics/molecular mechanics (QM/MM) methods

Combined quantum mechanics/molecular mechanics (QM/MM) methods allow the modelling of reactions within enzymes by combining the power and flexibility of a quantum chemical method with the simplicity of molecular mechanics. The fundamental basis behind the QM/MM approach is simple: a small part of the system is treated quantum mechanically, *i.e.* by an electronic structure method, for example at the *ab initio* or semiempirical molecular orbital, or density-functional theory QM level. The QM treatment allows the electronic rearrangements involved in the breaking and making of chemical bonds to be modelled. The QM region in a study of an enzymic reaction mechanism is the enzyme active site, *i.e.* the reacting groups of the enzyme, substrate and any cofactors. The large non-reactive part is described more simply by empirical molecular mechanics. Different types of coupling between the QM and MM regions are possible. For applications to enzymes, which are polar, it is important to include the interactions between the QM and MM regions. As noted above, modern molecular mechanics methods give a good description of protein structure and interactions, so can ensure that these are treated accurately. QM/MM calculations can be carried out at *ab initio*[82] or semiempirical[83] molecular orbital, density-functional[84] or approximate density functional (*e.g.* the self-consistent charge density functional tight-binding (SCC-DFTB)[85] method combines computational efficiency with reasonable accuracy for many applications) levels of QM electronic structure calculation.

The first enzyme to be studied with QM/MM methods was hen egg-white lysozyme, in a seminal study by Warshel and Levitt in the mid 1970s.[30] Interest in QM/MM methods has grown rapidly in recent years. Following many recent developments and applications, it is now clear that QM/MM calculations can provide useful insight into the mechanisms of enzyme-catalysed reactions.[75,86,87] Examples include identifying functions of active site residues (such as a conserved proline in two flavin-dependent monooxygenases[42,43]); investigations of mechanistic questions (*e.g.* comparing and differentiating between alternative proposed mechanisms)[20,21,23,25,27]

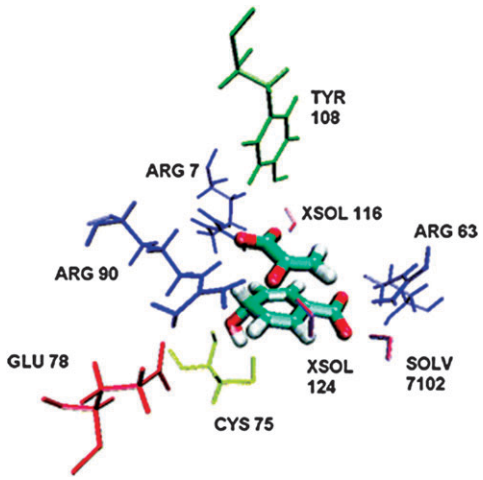



**Fig. 1** The transition state for reaction at the active site of *Bacillus subtilis* chorismate mutase from *ab initio* (RHF/6-31G(d)/CHARMM22) QM/MM modelling. Atoms treated by QM (the substrate/transition state) are shown in thick tubes; some important active site groups (treated by MM) are also shown (thin tubes). Reproduced from ref. 88.

and analysing catalytic contributions (such as the roles of conformational effects and transition state stabilization in chorismate mutase).[9,25,88] In this last case, a relatively simple enzyme reaction (a Claisen rearrangement), has given rise to wide debates on the origin of catalysis,[8,19] as discussed in Section 4.1 below. Modelling and simulation have been crucial in this case, and others, in formulating and testing mechanisms and hypotheses. The structure of the transition state found by *ab initio* QM/MM modelling of this enzymic reaction[88] is shown in Fig. 1.

Current extensions in the scope of the use of QM/MM methods include free energy perturbation simulations,[89] *e.g.* to calculate relative binding affinities, and in molecular docking and scoring of binding affinities.[90] QM/MM methods provide several advantages over MM methods in studies of ligands bound to proteins, including potentially a better physical description of a ligand (*e.g.* including electronic polarization), and avoiding the need for time-consuming MM parameterization for the ligand.

Many different QM/MM implementations are available, in several widely-used program packages. Transition state structures can be optimized.[91,92] Free energy differences such as activation free energies can be calculated, as can quantum effects such as tunnelling and zero-point corrections.[12,21,78] More approximate, less computer intensive, QM/MM methods (such as semiempirical or SCC-DFTB QM/MM) allow more extensive simulations to be performed (*e.g.* molecular dynamics or Monte Carlo simulations, extensive conformational sampling, or calculation of reaction pathways and Hessians). Specifically parameterized semiempirical methods can give improved accuracy for a particular reaction.[93,94] High-level QM/MM calculations (*e.g. ab initio* or density functional level QM) are required for some systems but can be extremely computationally demanding; they also have an important role in testing more approximate (*e.g.* semiempirical) QM/MM methods. The computational demands of

high level (*e.g. ab initio*) QM/MM calculations[38] typically limit their application to energy minimization/geometry optimization (as opposed to simulation/conformational sampling) to generate reaction paths, or 'single point' calculations on structures optimized at lower levels. This can lead to problems or misleading results, *e.g.* in enzyme reactions involving large movements of charge and strong interactions.[95]

One of the main differences between various QM/MM methods is the type of QM/MM coupling employed *i.e.* in how the interactions (if any) between the QM and MM systems are treated.[96] The simplest linking of QM and MM methods involves a straightforward 'mechanical' embedding of the QM region in the MM environment, where the interactions between the QM and MM regions are treated purely classically by MM, *i.e.* the QM system is represented by (MM) point charges in its interaction with the MM environment. In calculations of this type, the QM/MM energy of the whole system, $E_{\mathrm{TOTAL}}{}^{\mathrm{QM/MM}}$, is calculated in a simple subtractive scheme:

$$E_{\mathrm{TOTAL}}{}^{\mathrm{QM/MM}} = E_{\mathrm{TOTAL}}{}^{\mathrm{MM}} + E_{\mathrm{QM\ region}}{}^{\mathrm{QM}} - E_{\mathrm{QM\ region}}{}^{\mathrm{MM}} \quad (1)$$

where $E_{\mathrm{TOTAL}}{}^{\mathrm{MM}}$ is the MM energy of the whole system, $E_{\mathrm{QM\ region}}{}^{\mathrm{QM}}$, is the QM energy of the QM region and $E_{\mathrm{QM\ region}}{}^{\mathrm{MM}}$ is the MM energy of the isolated QM region. This simple subtractive approach can be applied to all combinations of theory levels (for example combining different levels of QM treatment (QM/QM) as opposed to QM with MM) and also forms the basis for the multi-layer ONIOM (Our own N-layered Integrated molecular Orbital and molecular Mechanics) method,[97] at least in its simplest form. A QM/QM calculation involves a high and a low level of QM theory, with a small region treated by a high level and the *entire* system treated at the low level (*e.g.* where the lower level theory is a semiempirical molecular orbital method such as AM1, PM3 or MNDO), polarization is included at the lower level of QM theory. An example of the application of this type of method to an enzyme reaction is an ONIOM-type study (with the MOZYME package) of the mechanism of citrate synthase.[98]

More intensive QM/MM methods include polarization of the QM region by the MM environment. This is likely to be important for many enzymes, given their polar nature. QM/MM methods of this type include electrostatic interactions between the QM and MM regions in the QM calculation. This models polarization of the QM system by the MM system, by directly including the MM atomic charges in the QM calculation. Leakage of charge from the QM region to the MM region has been found not to pose a significant problem in QM calculations.[99] The electronic structure calculation therefore includes the effects of the MM atoms. Methods of this type are the focus of the discussion here. A higher level of complexity yet would also in addition involve polarization of the MM region through the use of a polarizable MM force field, and potentially self-consistent polarization of the MM region through an iterative procedure. At present, models of this sort are computationally much more intensive and may not always yield better results.[100] Developing QM/MM methods of these types is a significant challenge partly because of the increased computational expense required for the calculation of polarization of the MM system. Also, most MM force fields that have been developed for biological macromolecules do not allow for polarization or indeed any changes in atomic charges. QM/MM methods that include polarization of the MM system have been developed for small molecular systems.[101] QM/MM calculations can assist in the development of polarizable MM force fields, for example in assessing polarization effects for small (QM) regions in large biomolecules.[99,102]

## 4. Examples of recent modelling studies of enzyme-catalysed reactions

When undertaking modelling of an enzyme-catalysed reaction, the choice of an appropriate method for the particular system and questions of interest is vital. A modelling method should be able to deliver a useful and reliable result in a reasonable time. Some strengths and weaknesses of current methods have been described above. To illustrate the capabilities of these methods in practice, some recent applications are discussed below. This field is still evolving, and it is not yet at the stage where quantitative, exact predictions of (for example) reaction rates or the effects of mutation can routinely be made. For this reason, it is important to validate predictions from modelling by comparisons with experimental data. For example, it can be useful to compare activation barriers for a series of alternative substrates with the activation energies derived from experimental rates: demonstration of a correlation can validate mechanistic calculations as being truly predictive.[43,103] Some enzymes have become important model systems in the development and testing of computational methods: comparison of results from different methods and protocols are important validations to provide confidence in modelling methods. Examples of such 'guinea pig' enzymes include chorismate mutase,[8,19,25,38] lysozyme,[2,30,104], citrate synthase,[4,33,98] $P450_{cam}$,[19,23,27] *para*-hydroxybenzoate hydroxylase,[27,38,42,43] triosephosphate isomerase[21,27,33] and methylamine dehydrogenase;[49,105] several of these are discussed below.

### 4.1 Chorismate mutase: analysing catalysis

Chorismate mutase illustrates well how modelling is contributing to, and helping to resolve, debates in enzymology. This enzyme is relatively simple, catalysing a single-step chemical transformation, but nonetheless is at the centre of controversies regarding the origin of its catalytic efficiency.[8,19,25] Chorismate mutase catalyses the Claisen rearrangement of chorismate to prephenate: this makes it an ideal system for analysing the root causes of catalysis, because the reaction does not involve any covalent interaction between the enzyme and the substrate, and also because the same reaction occurs in solution with the same reaction mechanism. The experimental free energy barrier $\Delta^{\ddagger}G$ = 15.4 kcal mol$^{-1}$ ($\Delta^{\ddagger}H$ = 12.7 kcal mol$^{-1}$) in *Bacillus subtilis* chorismate mutase is significantly lower than that for the uncatalysed reaction in aqueous solution ($\Delta^{\ddagger}G$ = 24.5 kcal mol$^{-1}$, $\Delta^{\ddagger}H$ = 20.7 kcal mol$^{-1}$).[106] This translates to a rate acceleration of $10^6$ by the enzyme ($\Delta\Delta^{\ddagger}G$ = 9.1 kcal mol$^{-1}$). QM/MM

calculations (*e.g.* at the semiempirical AM1/CHARMM or *ab initio* QM level) have previously shown TS stabilization by the enzyme.[25,81,88,107–111] This is also supported by mutagenesis experiments that show a significant decrease in catalytic activity when Arg90 is mutated to the isosteric but neutral citrulline residue.[112] The enzyme-bound conformation of chorismate is significantly different from that in solution, and more closely resembles the TS.[8,9,25,107,113–115] Bruice *et al.* have controversially argued that TS stabilization is not significant in chorismate mutase catalysis. Instead, these workers have proposed that catalysis is almost entirely due to the selection of a reactive conformation, described as a near-attack conformation (NAC).[7,116,117] This proposal, although it has mutated over time, has been vigorously promoted by these workers as vital not only in chorismate mutase, but as a potentially generally important effect in enzyme catalysis. It is similar in many respects to the old 'strain' hypothesis[118] that enzymes function by distorting their substrates into reactive conformations. One central problem with the NAC proposal, though, is that there is no unique definition of a NAC.

Bruice *et al.* based their estimates of NAC populations in chorismate mutase largely on unrestrained molecular dynamics simulations (*e.g.* in solution and in the enzyme). It is hence possible that high energy conformations will be sampled too infrequently (even in multi-nanosecond dynamics simulations), and thus the free energy cost for their formation may be overestimated.[8,9] This may lead to a significant overestimation of the catalytic benefit of NAC formation.

It is more useful to assess the catalytic benefit of forming the substrate conformation bound to the enzyme, rather than rely on a subjective definition of a 'NAC', or a definition derived by fitting to the observed catalytic effect. As noted by many workers, dating back to the first QM/MM study of the enzyme,[107] the conformation of chorismate bound to the enzyme is significantly altered from the conformation in solution (or indeed in the gas phase).[8,9,25,113,115] Fig. 2 depicts the distinct regions of phase space (in terms of C–C bond forming distance and attack angle) sampled by chorismate in solution and in the enzyme, from AM1/CHARMM MD simulations.[8] Potentially, the 'cost' of forming this conformation could be translated into a catalytic benefit. One can then ask what is the cost of forming the same conformation in solution as that when chorismate binds to the enzyme. Extensive free energy perturbation molecular dynamics methods give a free energy cost of 3.8–4.6 kcal mol$^{-1}$, or 5 kcal mol$^{-1}$ (by semiempirical QM/MM (AM1/CHARMM)[9] or empirical valence bond methods,[8] respectively), for forcing the chorismate in solution into the more restricted conformation found in the enzyme. This equates to a catalytic benefit of only around 40–55% of the total $\Delta\Delta^{\ddagger}G$ between enzyme and solvent. The good agreement between these findings, which applied completely different theoretical methods, is striking and suggests that this is a reliable estimate.

These results imply that catalysis cannot be due solely to binding of a reactive conformation, and therefore that stabilization of the TS (relative to the bound substrate) by the enzyme must be involved, in agreement with earlier QM/MM results for the enzyme-catalysed reaction. The

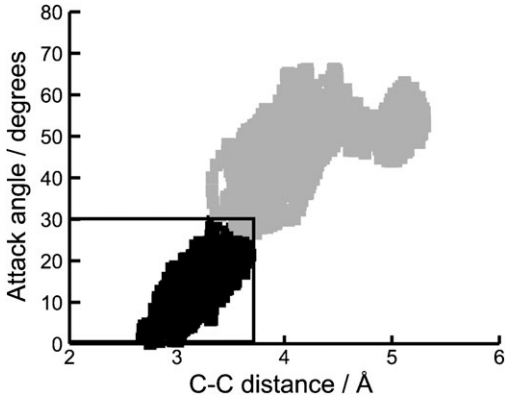



**Fig. 2** QM/MM (AM1/CHARMM) molecular dynamics simulations show that chorismate adopts a significantly different structure when bound to the enzyme chorismate mutase (black) than it does in solution (grey). This contributes to catalysis by the enzyme, but the effect is probably due to the enzyme's high affinity for the transition state of the chorismate to prephenate reaction. Reproduced from ref. 9.

reliability of these lower-level methods has been questioned, however. This key issue of whether the TS is stabilized relative to the bound substrate has been investigated with density functional theory QM/MM methods (B3LYP/6-31G(d)/CHARMM).[119] To study the effect of the enzyme on the reaction, 16 different QM/MM adiabatic reaction pathways were calculated using a combination[120] of the *Jaguar*[121] and *Tinker*[122] programs. The substrate was chosen as the QM region, treated at the hybrid density functional B3LYP/6-31G(d) level of theory, which gives a good description of this reaction.[38,81,88] The effects of including some amino acid side chains (*e.g.* Glu78 and Arg90) in the QM region are relatively small.[123,124] The QM/MM treatment of the active site interactions has been found to be accurate for chorismate mutase, because electrostatic interactions dominate.[110] The MM region was a 25 Å radius sphere of protein and solvent, treated with the CHARMM22 force field. All atoms in the outer 5 Å were held fixed, with all other atoms free to move. Starting structures were taken from semiempirical QM/MM (AM1/CHARMM22 and PM3/CHARMM22) molecular dynamics simulations of the TS. The difference in length between the forming C–C and breaking C–O bonds was used as a reaction coordinate: this has been shown to be a good choice for modelling the reaction.[25,81,88] Reaction pathways were calculated by restrained optimisations in both directions along the reaction coordinate. The average calculated barrier was 12.0 kcal mol$^{-1}$ (standard deviation, $\sigma$ = 1.7 kcal mol$^{-1}$), in excellent agreement with the experimental activation enthalpy (12.7 kcal mol$^{-1}$). At the TS, the average length of the breaking C–O bond was 2.02 Å, with a standard deviation, $\sigma$, of only 0.03 Å, while the average length of the forming C–C bond was 2.63 Å ($\sigma$ = 0.03 Å). A relatively large spread in the calculated energy barriers (9 to 15 kcal mol$^{-1}$) was observed, due almost entirely to differences in the protein environment.

Analysis of catalysis ideally requires comparison of energy profiles in the enzyme and in solution. The barrier in solution, relative to the enzyme-bound conformation, is similar to that in the gas phase,[88,110,113] so for this enzyme, the gas-phase

profiles can be taken as a convenient and meaningful reference. The difference between the gas-phase (QM-only) and QM/MM energy gives the stabilization of the reacting system by the protein environment. This term is large and negative along the whole reaction coordinate, because of favourable binding Coulombic interactions between the dianionic substrate and the positively charged side-chains in the active site. Most importantly, a systematic variation in the stabilization energy relative to the reactant complex was found along the reaction coordinate. In all cases, the TS was calculated to be stabilized significantly more than the reactant, while the product is generally destabilized (relative to the reactant). The stabilization of the TS is quite variable, and correlates very well with the computed barrier height. On average, the enzyme was found to stabilize the TS by 4.2 kcal $mol^{-1}$ more than it stabilizes the reactant. The TS stabilization is overwhelmingly electrostatic, in agreement with other calculations.[107,110] The correlation of barrier height with TS stabilization shows that the reactivity in the enzyme is determined by the degree of TS stabilization. It appears that conformational effects (*i.e.* binding of a reactive conformation),[25,114] and TS stabilization (relative to the bound substrate) contribute roughly equally to catalysis in this important model enzyme. The calculated average TS stabilization in this work (4.2 kcal $mol^{-1}$)[119] and the previously calculated cost of forming a reactive conformation in the enzyme, compared to solution[8,9] (3.8–5 kcal $mol^{-1}$) sum to give a value very close to the experimentally observed catalytic rate acceleration by chorismate mutase (experimentally, the difference in activation barriers for the reaction in solution compared to the enzyme is $\Delta\Delta^{\ddagger}G$ = 9.1 kcal $mol^{-1}$), suggesting that a combination of conformational effects and transition state stabilization together account for catalysis by the enzyme. The affinity of the enzyme is highest for the transition state, so both effects can be rationalized as arising from transition state stabilization by the enzyme.[8,119]

Chorismate mutase also provides an example of what is now possible with state-of-the-art QM/MM calculations. High level QM/MM calculations give activation energies in excellent agreement with experiments for chorismate mutase (and *para*-hydroxybenzoate hydroxylase). QM/MM calculations were carried out with QM methods including MP2 (Møller–Plesset second-order perturbation theory), LMP2 and coupled cluster (LCCSD(T0): coupled-cluster theory with single and double excitations; the L in the acronyms indicates that local approximations were used, and T0 is an approximate triples correction).[38] These *ab initio* electron correlation methods can calculate rate constants of gas-phase reactions involving only a few atoms with accuracy comparable to those in experiment. Theoretical developments[125,126] now allow their application in QM/MM calculations on reactions in enzymes. Only 'single point' energy calculations were possible using these highly computationally demanding methods; B3LYP/MM optimized reaction pathways provided the structures. Multiple pathways were generated to account for effects of conformational fluctuations. Only the LCCSD(T0) results were in close agreement with experiments for these enzymes. LMP2 and B3LYP barriers were too low by 3–5 kcal $mol^{-1}$, indicating that a high-level electron correlation treatment such as LCCSD(T0) is required for quantitative predictions of energy barriers in *ab initio* QM/MM calculations on enzymes.

Activation free energies were calculated by estimating activation entropies from low-level QM/MM umbrella sampling molecular dynamics simulations. The free energies were also found to agree exceedingly well with experiments for both enzymes. The sensitivity of the results to a variety of approximations was tested, an important feature in establishing confidence in the results. It is encouraging that these calculations were performed by different groups, using somewhat different QM/MM methods (*e.g.* differing in the MM force field applied), and yet both gave similar levels of agreement. Such QM/MM calculations remain extremely computationally demanding, but offer the prospect of quantitative predictions of enzyme mechanisms in first principles electronic structure calculations.

Transition state theory is the basis for the comparison between the calculations and experiments. The excellent agreement indicates that transition state theory describes these enzyme-catalysed reactions well. Dynamical effects apparently play only a small role in determining the reaction rates in the enzymes. Transition state theory is a good and sound basis for understanding such enzyme-catalysed reactions.

### 4.2 Hen egg-white lysozyme

Lysozymes are an important part of the chicken immune system, belonging to a large class of enzymes known as glycosidases. They catalyse the hydrolysis of the β(1–4) glycosidic bond between *N*-acetyl-muramic acid (NAM) and *N*-acetyl-D-glucosamine (NAG) in peptidoglycans, a component of bacterial cell walls. The prototypical lysozyme is that from hen egg-white. Hen egg-white lysosyme (HEWL) was the first enzyme to have its three-dimensional structure solved by X-ray crystallography.[127] It was also the first enzyme to have a mechanism proposed based on structural data and is hence one of the most important and widely studied enzymes in biochemistry. In this mechanism, proposed by Phillips *et al.*, a dissociative ($S_N1$-type) mechanism was hypothesised, in which Glu35 donates a proton to the glycosidic oxygen between rings D and E. This leads to breakage of the glycosidic bond and, according to Phillips *et al.*, formation of an oxocarbenium ion intermediate. Stabilization of the positive charge of this intermediate was proposed to be provided by the lone-pair of the NAM ring oxygen and the negatively charged carboxylate group of Asp52. A water molecule attacks the NAM $C_1$ on the only accessible side of the oxocarbenium ion, resulting in retention of the anomeric configuration. This mechanism is different to the general mechanism which was earlier proposed for configuration-retaining glycosidases by Koshland, in which the reaction instead proceeds *via* a covalent intermediate, following a $S_N2$-type mechanism, which also retains the original anomeric configuration (Fig. 3).[128] In this mechanism, a bond is formed between the D site NAM and Asp52. Experimental evidence for the Koshland-type (covalent) mechanism came much later, from crystallography and electrospray ionization mass spectrometry.[1,129] The experimental evidence is however based on the use of either mutated HEWL or unnatural substrates and hence does not

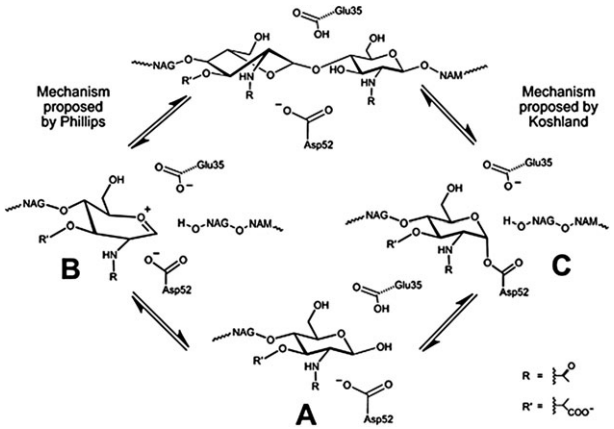



**Fig. 3** Catalytic mechanisms for lysozyme, proposed by Phillips and Koshland. QM/MM simulations by Bowman *et al.*[2] indicate that HEWL follows the Koshland mechanism. Reproduced from ref. 2.

categorically support the covalent pathway in the biologically relevant situation of wild type enzyme and natural substrate.

As mentioned above, lysozyme was first studied with a QM/MM method in 1976 by Warshel and Levitt.[30] They studied the stability of the oxocarbenium ion and concluded that electrostatic stabilization would be the most important factor in forming this suspected intermediate. In 1985 Post and Karplus carried out a (MM) molecular dynamics simulation of lysozyme with a NAG hexamer.[104] From their simulation they suggested a third mechanism in which an endocyclic bond is broken in the first step, without the requirement for substrate distortion. This mechanism had been proposed previously for some glycoside hydrolysis reactions, and has been shown to occur in some non-enzymic reactions. Density functional theory (*i.e.* QM only) calculations performed by Bottoni *et al.* on a small model system did not support the Phillips mechanism and suggested that both the Koshland-type and Post/Karplus mechanisms were energetically feasible.[130]

QM/MM MD simulations, with high-level energy corrections, have been used by Bowman *et al.* to investigate the mechanism of wild-type HEWL with its natural substrate.[2] Flexibility of the substrate and active site was accounted for by using QM/MM umbrella sampling simulations to generate free energy profiles. The crystal structure of the enzyme-product complex was used for simulations, which were run backwards from the product, simulating the second half of the reaction. The entire enzyme was modelled, solvated in a 25 Å radius sphere of explicit water molecules. The QM region comprised the entire D site NAM and side-chains of Glu35 and Asp52, and was treated with the semiempirical PM3 method. Link atoms were used to satisfy the valences of the atoms at the QM/MM boundary. The MM region was modelled with the CHARMM22 forcefield.[131] Reaction coordinates were chosen based on the breakage and formation of bonds during the reaction. The limitations of the PM3 method were overcome by calculating a correction to the PM3/CHARMM22 energy using higher levels of QM theory (both B3LYP/6-31+G(2d) and MP2/6-311+G(2d)). This correction was calculated by using a small model system consisting of acetic acid to model Glu35 and tetrahydropyran-2-ol to model the D site NAM. The two higher level methods were in good agreement for the energetics of proton transfer and revealed a signifcant overestimation of the barrier calculated with the PM3 method.

The initial protonation of the glycosidic oxygen by Glu35 was found to have a barrier of $\sim$16 kcal mol$^{-1}$ relative to the reactant complex, resulting in an intermediate lying 5 kcal mol$^{-1}$ higher in energy. The subsequent breakage of the glycosidic bond has a barrier of 12 kcal mol$^{-1}$ relative to this intermediate and is followed by spontaneous formation of the covalent intermediate analogous to that of the Koshland mechanism.[128] The transition state to breakage of the glycosidic bond was found to display oxocarbenium ion character. The total barrier to formation of the covalent intermediate is $\sim$18 kcal mol$^{-1}$, which is in good agreement with the value of 17 kcal mol$^{-1}$ derived from the experimental rate constant for the forward reaction.[132] Although the reverse reaction was modelled in this study, the barrier is likely to be similar to that of the forward reaction. The stability of the oxocarbenium ion and covalent intermediate was compared by simulating their interconversion. The free energy barrier showed the covalent intermediate to be significantly more stable, by around 30 kcal mol$^{-1}$. The endocyclic mechanism proposed by Post and Karplus[104] was also simulated, but found to be prohibitively high in energy compared with the exocyclic protonation mechanism.

The calculations performed by Bowman *et al.* provide convincing evidence that reaction in HEWL proceeds *via* a covalent intermediate. This finding, which agrees with experimental findings for mutant HEWL and/or unnatural substrates,[1] is contrary to the mechanism shown in most biochemistry textbooks.

### 4.3 Citrate synthase

Citrate synthase is an important model for mechanisms of carbon–carbon bond formation in biological catalysis and has been extensively studied both experimentally and computationally.[4,98,133–139] It catalyses the first step of the citric acid cycle, the conversion of oxaloacetate to citrate using acetyl-coenzyme A (acetyl CoA). The enzyme-catalysed reaction starts with enolization, where a proton of acetyl-CoA is abstracted by an aspartate residue. The subsequent condensation step is the nucleophilic attack on the carbonyl carbon of oxaloacetate by the enolate intermediate. This produces citryl-CoA which is converted to citrate and CoA upon hydrolysis. (Fig. 4) Citrate synthase is another example of an enzyme reaction mechanism given in many textbooks that is probably incorrect in important details. Computational enzymology has been central in determining the mechanism of the enolization[4,136,140,141] and, more recently, the condensation step for this enzyme.[142] In the condensation step, formation of citryl-CoA requires protonation of what was the carbonyl oxygen of oxaloacetate. Several possible mechanisms have been put forward. These included one of the several histidines in the active site acting as a proton donor,[133,134,139] or protonation occurring with hydrolysis.[137] A further possibility is that an arginine residue, which was found to form a hydrogen bond to the oxaloacetate carbonyl oxygen, might donate the proton.[133] This latter hypothesis was deemed somewhat controversial given that acid/base catalysis

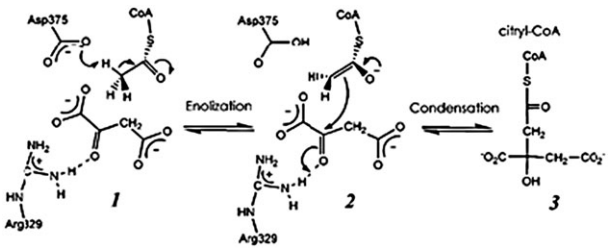



**Fig. 4** Formation of citryl-CoA in citrate synthase. High level QM/MM calculations suggest that an arginine residue acts as the acid in the condensation step. Reproduced from ref. 142.



by arginine residues is rarely observed, and arginine is intrinsically highly basic.

High level *ab initio* QM/MM methods were used by van der Kamp *et al.* to model the condensation reaction in CS, revealing that the arginine can act as a proton donor, leading to a stable citryl-CoA intermediate.[142] Calculations were performed on a structure of chicken citrate synthase co-crystallized with acetyl-CoA and *R*-malate (PDB code 4CSC), initially modelling the enolization and condensation steps at the AM1/CHARMM27 level. Several mechanisms for condensation were modelled but the mechanism with the arginine acting as a proton donor gave the most favourable results. Starting from the AM1/CHARMM27 optimized structure for the transition state for enolization, the energy profile for the entire reaction was optimized at the higher B3LYP/6-31+G(d)/CHARMM27 level, using QoMMMa.[120] The QM region consisted of the thioester part of acetyl-CoA, oxaloacetate, the Arg329 side chain and the Asp375 and Asp327 side chains. High-level *ab initio* single point energies were calculated resulting in a profile at the SCS-MP2/aug-cc-pVDZ//B3LYP/6-31+G(d)/CHARMM27 level. The energy barriers obtained for enolization and condensation were 10.2 kcal $mol^{-1}$ and 14.2 kcal $mol^{-1}$, respectively, and indicate that the enolization and condensation steps are closely linked and have similar barriers. The use of a variety of reaction coordinates for the condensation step revealed that carbon–carbon bond formation is concerted with proton transfer from the arginine to the former oxaloacetate carbonyl oxygen. The calculated overall reaction energy barrier of 14.2 kcal $mol^{-1}$ is consistent with the barrier of 14.7 kcal $mol^{-1}$ derived from the experimental overall reaction rate.[143,144]

The high level *ab initio* QM/MM calculations reveal an unexpected mechanism for this 'textbook' enzyme. The results are consistent with, and shed new light on experimental studies. The implications are significant, as the role of arginine as an acid/base in enzyme-catalysed reactions may be more widespread than previously anticipated.

Other aspects of citrate synthase catalysis have also been examined computationally. For example, QM/MM methods have been used to investigate the contribution of substrate polarization in catalysis. Experimental spectroscopic measurements, by Fourier transform infrared spectroscopy (FTIR)[145] and NMR,[146] have shown that the oxaloacetate carbonyl is polarized when bound to the enzyme. This has been proposed to be an important catalytic effect, of potential general importance: many substrates are observed to be polarized when bound to enzymes. Modelling using QM/MM techniques, with different degrees of substrate polarization, showed that the effect of substrate polarization on the reaction barrier is small.[94] Treating oxaloacetate by MM, the polarization of the carbonyl was varied in a computational experiment, to test systematically the effect on the reaction. This showed that the effect of polarization is quite small, much smaller than that of electrostatic stabilization in the active site. This provides further support to the picture that catalysis in citrate synthase is primarily due to the stabilization of the acetyl-CoA enolate. In psychrophilic organisms, the CS enzyme tends to be less thermally stable, in order to counteract freezing. Enolate formation has been modelled in psychrophilic CS using the EVB method.[139] A lower activation enthalpy and entropy was observed in the psychophilic CS compared with mesophilic and hypothermophilic CS. These observations are linked to a decreased protein stiffness in the psychophilic CS, which extends beyond the active site.

### 4.4 Cytochrome P450

Cytochrome P450s form an ubiquitous class of haem enzymes, which function as mono-oxygenases in steroid biosynthesis and a wide range of biological reactions, including the metabolism of potentially toxic hydrocarbons. These enzymes are of significant pharmaceutical interest, because of their roles in drug metabolism.[20,23,27,103] Better understanding of the mechanisms of P450 enzymes will help in predicting biotransformations of pharmaceuticals and other xenobiotics, and should therefore assist drug development.[20] Modelling is playing a central role in analysing the mechanism and determinants of specificity in reactions catalysed by cytochrome P450. A large number of cytochrome P450 proteins are known, with widely varying specificities and activities. The active haem is an iron protoporphyrin IX complex, linked covalently to a cysteine residue in the protein. The catalytically active form of the enzyme for oxidation is thought to be a haem oxoiron (IV) porphyrin radical cation, the so-called Compound I.[147] Compound I is highly reactive, which prevents its experimental isolation and characterization. Among the many reactions catalysed by P450 enzymes are the hydroxylation of alkanes and aromatic compounds, and the epoxidation of alkenes. Many of these reactions are potentially useful in synthetic and other practical applications.

Shaik and co-workers have used density functional theory and QM/MM techniques to examine various aspects of the mechanism of alkane hydroxylation by cytochrome P450.[148–151] This has included calculation of the potential energy surface for the so-called 'rebound' mechanism with methane as a substrate. In the rebound mechanism, Compound I initially abstracts a hydrogen atom from the alkane. This step is followed by recombination of the hydroxo-radical on the iron with the alkyl radical, generating the ferric-alcohol complex. Calculations were carried out on a model of Compound I, with $SH^-$ used to represent the $SCys^-$ ligand. Calculations were performed at the B3LYP hybrid density functional theory level using the *Gaussian*[152] and *Jaguar*[121] programs, using the Los Alamos effective core potential (ECP) coupled with the double-ξ LACVP basis set

for iron and a 6-31G basis for all other atoms. Compound I contains 3 unpaired electrons, giving rise to two closely lying spin states, a doublet (low spin) and a quartet (high spin). For the first step of the reaction, the two spin states showed nearly identical reactivity (a reaction barrier in both cases of approximately 27 kcal $mol^{-1}$). However, for the second step of the rebound mechanism, the high spin state was found to have a barrier of approximately 5 kcal $mol^{-1}$ while there was no barrier to reaction for the low-spin state. With ethane as a substrate, the two spin states were found to be nearly degenerate with a barrier to the second step of approximately 5 kcal $mol^{-1}$.[153]

Shaik and co-workers have also studied the mechanism of ethene epoxidation by Compound I[154,155] with similar techniques. They found that the barrier for the C–O bond formation leading to the radical intermediate was 14–15 kcal $mol^{-1}$ from either the high or low spin state of Compound I. The second step, to form the epoxide, was found to proceed from the low spin state with no barrier, and from the high spin state with a barrier of approximately 3 kcal $mol^{-1}$. These results indicate that the reaction of the low spin state is effectively concerted while that of the high spin state reaction is stepwise. The concerted reaction in the high spin state was also found to have a barrier only around 4 kcal $mol^{-1}$ higher than the stepwise reaction. The possibility of intermediates with differing electronic configurations, and significantly different lifetimes, may help to explain experimental observations such as occasional *cis*/*trans* isomerization and the production of aldehydes.

Research on aromatic hydroxylation by cytochrome P450 provides an example of how quantum chemical calculations on small models can help in developing structure-reactivity relationships. Hydroxylation of C–H bonds is a particularly important class of reaction in drug metabolism,[156] which can activate prodrugs, or affect the bioavailability of pharmaceuticals. For the reliable prediction of pharmaceutical metabolism and toxicology (ADME/TOX) properties, a key aim is the development of structure–activity relationships to predict conversions of drugs. Structure–activity relationships based on the structures and properties of substrates alone have been found to be of limited use. More detailed models are needed (most likely including information obtained from QM calculations), which include effects of the specificity of different cytochrome P450 isozymes.

Bathelt *et al.*[157,158] (Fig. 5) investigated hydroxylation of simple aromatic compounds by Compound I, in a model consisting of the porphyrin (without side chains), and the cysteinate iron ligand represented by a methyl mercaptide group ($CH_3S^-$). For the addition of Compound I to benzene, two different possible orientations of the substrate approach were found ('side on' and 'face on', the first with a lower barrier). Both orientations (Fig. 5) may be important in the reactions of different drugs in different P450s. The transition state for aromatic hydroxylation was found to have mixed radical and cationic character. This insight from calculations led to the development of new structure-reactivity relationships for substituted aromatics, using a dual-parameter approach combining radical and cationic electronic descriptors.

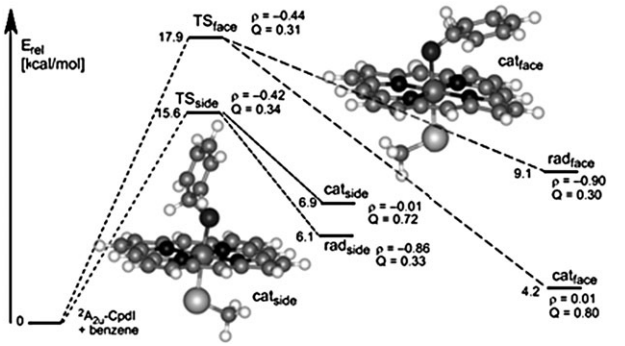



**Fig. 5** B3LYP potential energy surfaces for face-on and side-on addition of benzene to the Compound I model species, with the latter having the lower barrier. Energies of both the cation- and radical-like adducts are shown. $\rho$ and Q are Mulliken charge and spin densities on the substrate, respectively. Reproduced from ref. 158.

Different P450 isoenzymes show very different substrate specificity and hydroxylation patterns. These could be the result of orientation or binding effects,[159] or the intrinsic chemical reactivity of different positions in the substrates.[160] Genetic polymorphisms can also have significant effects, *e.g.* in determining drug metabolism.[161] It is possible also that the electronic properties of Compound I could be modulated by the protein environment,[147] and that this could be a key factor in determining the reactivity of cytochrome P450s. To investigate questions of this sort, methods (*e.g.* QM/MM calculations) that include the protein explicitly are needed. QM/MM modelling of human cytochrome P450 enzymes (including complexes with the drugs diclofenac and ibuprofen) demonstrates the potential for QM/MM methods to contribute directly to practical questions of drug metabolism.[20,162]

Bathelt *et al.*[163] have also modelled the hydroxylation of benzene in the enzyme environment of CYP2C9, using QM/MM methods. In contrast to the gas-phase model calculations, the side-on and face-on pathways were found to have similar barriers, indicating that they might compete in reaction. The calculated QM/MM barriers were found to be consistent with the experimental rate constant for benzene hydroxylation in CYP2E1.[164] The rearrangement pathways from the initial σ-complex were modelled, to form the epoxide, ketone and N-protonated porphyrin species. Epoxide and ketone products were formed with ease in the face on pathway, whereas the epoxide product was favoured in the side on pathway. The results conclude that several pathways are energetically possible during P450 mediated aromatic hydroxylation.

QM/MM studies of bacterial $P450_{cam}$ have raised controversial issues about the factors important in the activity of the enzyme.[165,166] $P450_{cam}$ is a bacterial enzyme that catalyses the 5-*exo*-hydroxylation of camphor. It was the first cytochrome P450 enzyme to be crystallized and has been the subject of many subsequent experimental and computational studies. A large number of crystal structures of this enzyme are available, including several intermediates of the catalytic cycle and their mutants. The mechanism of the reaction has been studied with quantum chemical and QM/MM methods,[148] *e.g.* to characterize the active species,[167] and spectroscopic

properties.[168] Computational investigations of the hydroxylation of camphor,[165,166] the protein resting state[169] and the enzyme-product complex,[170] and the full catalytic cycle of the enzyme[171] have also been carried out. Recently Thiel and co-workers investigated the effect of mutation of Thr252 on proton delivery to the active site.[172]

Some calculations have raised a number of controversial issues concerning the nature of the active species and the reaction mechanism, for example calculations suggested that a propionate side chain on the haem group in the active species becomes partially oxidized, and carries unpaired electron density.[166,171] This effect was proposed to play an important role in the enzyme's catalytic properties. This was not, however, confirmed by other studies.[167,169] Different investigations also produced somewhat different barrier heights for camphor oxidation. Groups have worked together to try to reconcile these conflicting findings,[173] but controversy still remains.[174,175]Argument centres in particular on the possible role of haem propionates in catalysis. A central question is whether there is unpaired electron density on these side chains of haem. Żurek *et al.* calculated QM/MM reaction pathways for hydrogen abstraction from camphor in $P450_{cam}$, and carried out (MM) molecular dynamics simulations of the Compound I enzyme-substrate complex.[176] They found that this spin density is only found on the haem propionate when a charged aspartate is in close proximity, and is significant only when the unrelaxed crystal structure of the protein is used—geometry optimization and molecular dynamics equilibration lead to structures with reduced spin density. A much better model is obtained when Asp297 is treated as protonated (neutral): this preserves the protein conformation around the haem propionates during MD equilibration and QM/MM optimization, and shows no spin density on haem propionates. This agrees with other findings, produced independently with different QM/MM methods.[167,175] This shows that QM/MM results are reproducible despite the complexity of these metalloenzyme systems. The effects caused by the charged aspartate are a reminder of the need to consider alternative protonation states in proteins, and the value of molecular dynamics simulations in complementing reaction modelling. The barrier calculated by Żurek *et al.* was approximately 18 kcal $mol^{-1}$, which will be reduced by zero point energy (and perhaps some tunnelling) to an effective barrier of around 14 kcal $mol^{-1}$, consistent with rapid hydrogen atom abstraction by Compound I.

QM/MM methods have been used to show that the chemoselectivity of alkene oxidation in $P450_{cam}$ is dictated by the reactivity of Compound I.[177] The oxidation of cyclohexene by $P450_{cam}$ has been shown experimentally to result in the formation of a mixture of hydroxylation and epoxidation products, in contrast to propene oxidation by $P450_{LM2}$, where only epoxidation products are observed. In small alkenes such as these, the accessibility of the respective functional groups does not have a significant effect on the ratio of epoxidation and hydroxylation products. This was confirmed by MD simulations, which displayed free-tumbling of both alkenes in the active site of $P450_{cam}$, with significant time spent in positions where both epoxidation and hydroxylation could occur. The QM/MM (B3LYP/CHARMM) calculated energy barriers to epoxidation and hydroxylation of cyclohexene and propene are consistent with the experimentally observed selectivity. This work shows QM/MM to be a promising method in the prediction of chemoselectivity in drug metabolism where Compound I reactivity is a governing factor.

### 4.5 Aromatic amine dehydrogenase and methylamine dehydrogenase: enzyme dynamics and quantum tunnelling

The role of quantum tunnelling in enzyme-catalysed proton, hydride and hydrogen atom transfers is a topic of lively current debate in enzymology.[11–14,16,21,22,48,49,61,62] Experimental studies of kinetic isotope effects (KIEs) have demonstrated the role of quantum mechanical tunneling in enzymic H-transfer reactions.[12,15,48,49,178,179] In some cases kinetic isotope effects (KIEs) found on substitution of hydrogen by deuterium or tritium are very large, and in some cases show unusual temperature dependence. There has been considerable debate on the contribution of tunnelling to enzyme catalysis, *i.e.* whether tunnelling is enhanced in the enzyme environment compared to an equivalent reference reaction in solution. The possible role of protein dynamics in 'driving' tunnelling has been widely and hotly debated. Molecular simulations allow the investigation of these important questions.

Aromatic amine dehydrogenase (AADH) and methylamine dehydrogenase (MADH) are bacterial quinoprotein enzymes (dependent on a tryptophan tryptophyl quinone (TTQ) cofactor) that share 30% sequence identity, and catalyse the oxidative conversion of aromatic and aliphatic primary amines, respectively, to the corresponding aldehyde and ammonia:[49]

$$RCH_2NH_2 + H_2O \rightarrow RCHO + NH_3 + 2H^+ + 2e^-$$

The rate-limiting step in the reductive half-reaction involves proton transfer (breaking a C–H bond) from the substrate-derived iminoquinone intermediate to a catalytic aspartate.[179] Experimental and computational studies show a significant contribution from quantum tunnelling in the oxidative deamination of tryptamine by AADH.[12] The experimentally observed H/D kinetic isotope effect (KIE) of $55 \pm 6$ [12] is one of the largest yet found for an enzyme reaction, and is significantly above the semi-classical limit of around 7. MADH also shows KIEs higher than this semi-classical limit for the reaction with methylamine and ethanolamine, though lower than those for AADH with tryptamine.

Both AADH and MADH have been studied by QM/MM modelling using variational transition state theory with the small curvature approximation for tunnelling corrections (VTST/SCT). These VTST/SCT methods[78] allow KIEs to be calculated for enzyme reactions, and thus comparison with experiments. However, it is important to remember that the complexity of multi-step enzyme reactions can make direct comparison difficult.

There has been considerable controversy about the possible role of protein dynamics in enzyme reactions, particularly those involving quantum tunnelling. Protein dynamics could potentially affect enzyme reactions in two distinct ways. First, the population of reactive enzyme complexes may govern the overall reaction rate: reaction may proceed preferentially *via* a more reactive conformation, even if this is only a minor and

transient conformation. This is clearly an equilibrium effect, and both NMR[47] and simulations[77] show that this may be an important factor in some (perhaps many) enzymes. On the other hand, one could conceive of a coupling between protein dynamics and the reactive event, *i.e.* dynamics 'driving' reaction, although there are large differences in timescales between protein dynamics and those of breaking and making chemical bonds. Where this latter type of dynamic effect has been analysed, it has been found to play only a small part in determining reaction rates in enzymes, and probably even a smaller catalytic role (*i.e.* similar motions may be expected in other environments, *e.g.* in solution).[21,22] The results for AADH show no role for large-scale protein dynamics in driving reaction.[12] First, the VTST/SCT calculations are carried out with the (large, MM part of the) protein frozen: despite this, large kinetic isotope effects and large tunnelling contributions are found. These are sensitive to the exact configuration selected (from a molecular dynamics simulation), with some configurations favouring reaction (and tunnelling) more. Good results are obtained, however, with no inclusion of the effects of protein motion on the reaction itself. Crystal structures of the enzyme at various stages in its catalytic cycle show no evidence of large scale structural changes during reaction.[12] Finally, analysis of classical molecular dynamics simulations shows that the motion of the two crucial groups involved in proton transfer (Asp128 and the substrate) is apparently coupled only to a directly hydrogen bonded group in the active site: the simulations show no evidence of coupling over longer ranges, nor of networks of coupled motion associated with 'driving' reaction.[12] While none of these results constitutes proof that complex protein dynamics does not drive the reaction, they imply that the effect of dynamics is small.

MADH has been the subject of several computational modelling studies.[180–184] For example, Alhambra *et al.* carried out VTST/SCT QM/MM calculations of kinetic isotope effects for the proton transfer step in the methylamine-to-formaldehyde reaction in this enzyme.[185] The MM region was treated with the CHARMM22 MM force field, and consisted of 7248 protein atoms and 1243 water molecules. The semi-empirical method PM3 was used for the QM region, but with parameters specifically optimized for this reaction. The two regions were linked by means of the GHO QM/MM partitioning method.[186] Umbrella sampling molecular dynamics simulations were used to calculate the classical potential of mean force along an approximate reaction coordinate, defined as the difference of two bond lengths: the breaking C–H bond, and the forming O–H bond. The difference between the quantal and classical vibrational free energies along the path was calculated for the reactants and the variationally optimized transition state.[187] This procedure gives a quantum-corrected potential of mean force.[188] The transmission coefficient, including the effects of tunnelling, was then calculated. The classical activation free energy was calculated as 20.3 kcal mol$^{-1}$. When quantum mechanical vibrational energy was included, the barrier height was reduced to 17.1 kcal mol$^{-1}$. Including quantum mechanical tunnelling contributions gave an effective (phenomological) activation energy of 14.6 kcal mol$^{-1}$, a finding which can be compared directly with experiment, and agrees well with the experimental value of around 14.2–14.4 kcal mol$^{-1}$. A hydrogen/deuterium primary KIE of 18.3 was calculated for the perdeuterated substrate, agreeing well with the experimental value of 17.2. Without tunnelling, the calculated KIE was only 5.9. The VTST/SCT approach allows the various different contributions (for example, of tunnelling) to be calculated separately, unlike some other methods used to incorporate quantum effects.[189] These results demonstrate the central role of proton tunnelling in the reaction catalysed by MADH.

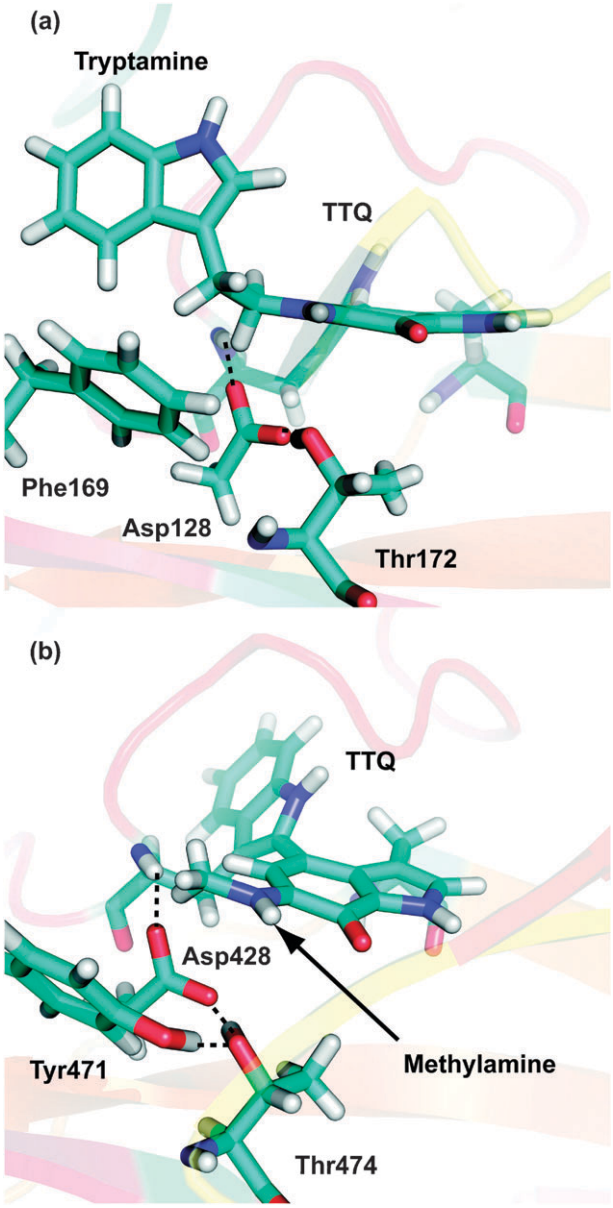



**Fig. 6** QM/MM optimized geometry of the active site of the (a) AADH/tryptamine and (b) MADH/methylamine (MA) enzyme-substrate complexes, from PM3-CHARMM22 MD simulations. Simulations show that quantum tunnelling is important in these enzymes. Different hydrogen bonding environments (dotted lines) of the catalytic aspartate are a possible cause of the difference in kinetic isotope effects observed between AADH and MADH.[105]

Tresadern *et al.* examined proton transfer in MADH for two substrates (methylamine and ethanolamine),[190] also with QM/MM VTST/SCT methods. These calculations used PM3 with their own specific reaction parameters (with the Gaussian94 program[152]), with the MM system described by the AMBER force field. The two regions were connected using the link atom approach. The calculated kinetic isotope effects were close to the experimental values, though covering a wide range. Two different modes were found to be possible for the ethanolamine substrate, giving rise to quite different kinetic behaviours. One configuration was found to give rise to considerably more tunnelling than the other. It was suggested that this conformational behaviour may be the cause of the different temperature dependence found experimentally for the KIEs of the two substrates.

Recently a wavefunction propagation approach similar to that developed by Warshel and co-workers[191] has been used to calculate the primary H/D kinetic isotope effect for the rate-limiting proton transfer in the deamination of methylamine in MADH.[192] Two different approaches were tested: one where the quantum particle moves on the potential created by the classical environment, but the classical environment evolves independently of the quantum particle and an approach where the quantum and classical regions evolve together using a mixed quantum/classical propagation scheme. The first approach gave a KIE of 16.2 ± 1.4 for the proton transfer step in the MADH/methylamine system and the second gave a KIE of 14.6 ± 1.8, in good agreement with each other and the experimental value of 16.8 ± 0.5.[193]

Ranaghan *et al.*[105] have analysed the key deprotonation step of the reaction of methylamine in MADH, and compared this reaction with that of AADH with tryptamine. They used QM/MM techniques, in classical molecular dynamics simulations and VTST calculations with multidimensional tunnelling corrections averaged over an ensemble of paths. The results revealed potentially important mechanistic complexity, namely two possible proton transfer routes—to either of the carboxylate oxygens of the catalytic base (Asp428). The activation barriers and tunnelling contributions for the two possible proton transfers were found to be similar, with a phenomenological activation free energy of 16.5 ± 0.9 kcal $mol^{-1}$ for transfer to either oxygen (PM3/CHARMM calculations, applying PM3-SRP specific reaction parameters[185])—in good agreement with the experimental value of 14.4 kcal $mol^{-1}$. In contrast, for the AADH system, transfer to the equivalent OD1 was found to be preferred. The hydrogen bond of Thr474(MADH)/ Thr172(AADH) to the catalytic carboxylate, and in MADH the non-conserved active site residue Tyr471 (equivalent to Phe169 in AADH) are identified as important factors in determining the preferred oxygen acceptor (see Fig. 6). The protein environment significantly affects the reaction energetics, and hence tunnelling contributions and KIEs. These environmental effects, and the related clearly different preferences for the two carboxylate oxygen atoms (with different KIEs) in MADH/MA and AADH/tryptamine, are possible causes of the differences observed in the KIEs between these two important enzyme reactions.[105,194]

### 4.6 Fatty acid amide hydrolase

QM/MM calculations have identified an unusual mechanism in fatty acid amide hydrolase (FAAH),[195] an enzyme involved in endocannabinoid metabolism, and a promising target for treatment of central and peripheral nervous system disorders. It is responsible for hydrolysing and thus deactivating bioactive lipid amides including the endocannabinoid anandamide. Lodola *et al.* calculated potential energy surfaces for the first step of the acylation reaction of FAAH with oleamide, using PM3/CHARMM QM/MM calculations. Higher level corrections, with hybrid density functional theory, were applied to overcome the limitations of the PM3 semiempirical method for the reaction energetics. The B3LYP/6-31 + G(d)//PM3-CHARMM results, which give a calculated barrier consistent with experiments, indicate that neutral Lys142 deprotonates the nucleophilic residue Ser241 *via* Ser217, which acts as a proton shuttle during the reaction (Fig. 7). This mechanism is supported by QM/MM free-energy perturbation calculations and Monte Carlo simulations at the PDDG/PM3 QM level, performed by Tubert-Brohman *et al.*[196] Simulations at this level by these workers of the reaction in the Lys142Ala mutant reproduce the experimental observation that this mutation decreases the rate of hydrolysis for oleamide significantly more than for methyl oleate. Simulations have also highlighted interesting conformational effects in (wild type) FAAH: QM/MM and molecular dynamics simulations of fatty acid amide hydrolase (FAAH) show that reaction (acylation with oleamide) occurs *via* a high energy conformation of the enzyme.[77]

QM/MM modelling has also been used to study the mechanism of a covalent (*O*-arylcarbamate) inhibitor binding to the enzyme, and its mode of binding.[197] Carbamic acid aryl esters, such as URB524 and its derivatives, are pharmaceutical lead compounds: they inhibit FAAH irreversibly, and in rodents show analgesic, antidepressant and anxiolytic type effects.[198] There has been uncertainty over the mode of binding of these inhibitors: they can be docked in two possible orientations in the FAAH active site; both modes place the carbamic group close to the nucleophilic Ser241.

QM/MM modelling of covalent adduct formation for URB524 in FAAH (using techniques similar to those used in modelling the reaction with the substrate oleamide described above[195]) shows that only one binding orientation is consistent with the experimentally observed irreversible carbamoylation of the nucleophilic serine. This is potentially crucial insight for designing new covalent inhibitors of this promising drug target, and understanding structure–activity relationships for these inhibitors. This application is an example of how QM/MM modelling can give useful insight for drug design (*e.g.* for covalent inhibitors), when traditional docking approaches alone may fail.

### 4.7 Overview of other recent modelling studies of enzyme mechanisms

The number of modelling studies of enzymes and their reactions continues to increase.[35,87] Only a few studies can be mentioned here. Interesting and representative studies include density functional modelling of the mechanisms

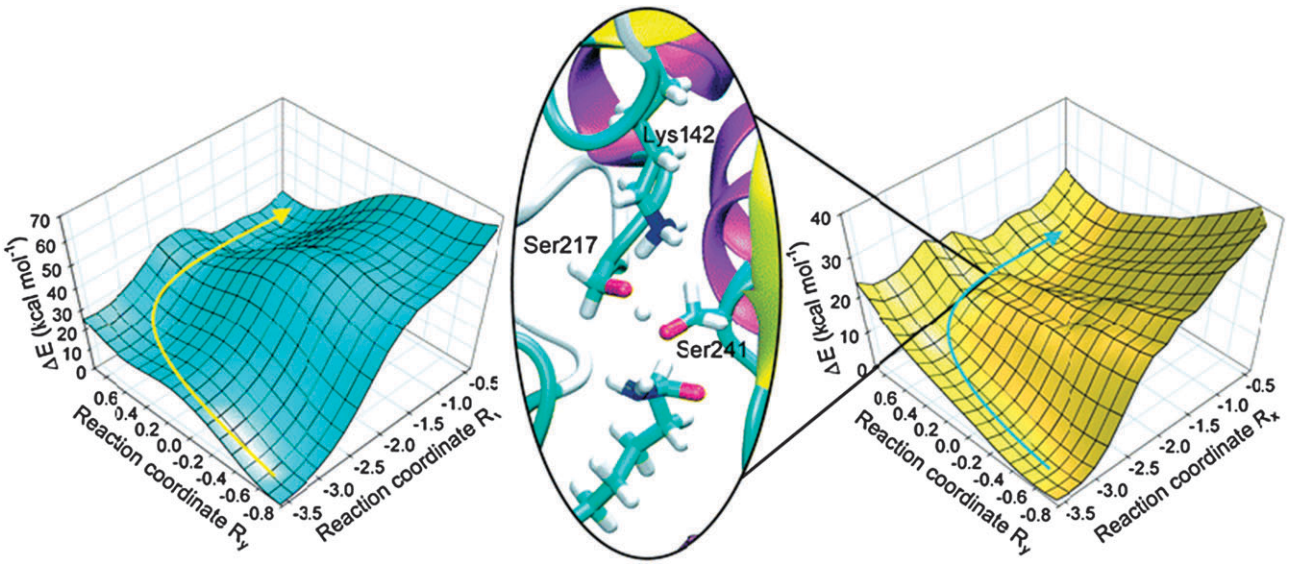



**Fig. 7** QM/MM potential energy surfaces for the formation of the tetrahedral intermediate for acylation of oleamide in FAAH. Left (in cyan), the potential energy surface calculated at the PM3/CHARMM22 level and right (in yellow), the B3LYP//PM3/CHARMM22 surface. Arrows indicate approximate reaction paths. In the middle, the transition state structure (as identified from the B3LYP QM/MM surface) is shown, showing proton transfer from Ser241 to Ser217 and from Ser217 to Lys142. QM/MM modelling also revealed complex conformational effects in the reaction.[77] Reproduced from ref. 211.

of naphthalene dioxygenase,[199] class III ribonucleotide reductase[200] and 4-hydroxyphenylpyruvate dioxygenase.[201] PM3/CHARMM QM/MM methods have been used to model the formation of the Meisenheimer intermediate in 4-chlorobenzoyl-CoA dehalogenase.[202] Proline isomerization in cylcophilin (and mutant proteins) has been investigated with SCC-DFTB/CHARMM QM/MM methods.[203] QM/MM methods have been used to calculate kinetic isotope effects in chorismate mutase[204] and catechol *O*-methyltransferase.[205] A combination of QM/MM molecular dynamics simulations and density-functional calculations have been applied to study a metallo beta-lactamase.[206,207] In 4-oxalocrotonate tautomerase, QM/MM methods have been used to investigate the contribution of the protein backbone in the mechanism.[208] QM/MM methods have been used to study the factors that control proton transfer in carbonic anhydrase.[209] The oxidation of acetaldehyde by aldehyde oxidoreductase, a mononuclear molybdenum enzyme, has also been modelled with QM/MM.[210] Multiple steered molecular dynamics simulations with a density-functional QM/MM technique have been used to calculate the free energy profile in chorismate mutase.[212] QM/MM molecular dynamics simulations of wild-type and mutant xylanases have revealed the role of an active site tyrosine in the stabilization of a boat conformation of the substrate.[213] QM/MM Monte Carlo free energy perturbation simulations have been applied to study the mechanism of macrophomate synthase, comparing the Diels–Alder with the Michael–Aldol reaction mechanism.[214] An interesting QM/MM study modelled inhibition mechanisms of neutrophil elastase by peptidyl alpha-ketoheterocyclic inhibitors of human neutrophil elastase, and highlighted the potential of QM/MM calculations in structure-based drug design.[215] The stereoselectivity involved in the inhibition of cysteine proteases by epoxides has been modelled for the first time using QM/MM.[216] Dinner *et al.* demonstrated substrate autocatalysis in uracil DNA-glysosylase by QM/MM modelling.[217] Simulations have examined the nature of the proton bottleneck in redox-coupled proton transfer in cytochrome c oxidase.[218] Many computational investigations,[16] in addition to those described above, have focused on quantum tunnelling in enzyme-catalysed proton, hydride or hydrogen atom transfer reactions, including *e.g.* studies of lipoxygenase.[219–221] The electronic structure of Compound I in cytochrome C peroxidase and ascorbate peroxidise has been calculated by QM/MM methods,[222] and compared with that in cytochrome P450.[147] The catalytic cycle of nitric oxide synthase, another haem-containing enzyme, has been studied with DFT and QM/MM methods and a novel mechanism of oxygen activation has been established.[223]

Bjelic and Åqvist used a well-tested homology model to examine the substrate binding mode and reaction mechanism of a malarial protease with a novel active site.[41] This enzyme (histo-aspartic protease (HAP) from the malaria parasite *P. falciparum*) is a target for antimalarial drug design, but its three-dimensional structure is not yet known. These workers used a combination of homology modelling, automated docking, and molecular dynamics/reaction free energy profile simulations to predict the structure of the enzyme and the conformation of the bound substrate. Finally, these calculations were used to predict the mechanism of the enzymic reaction.[41] The only amino acid residue found to be involved directly in the reaction was a catalytic aspartate, with stabilization by a histidine residue. The calculated reaction rate agreed well with experimental kinetic data for a hexapeptide substrate derived from human haemoglobin.

Specifically parameterized semiempirical QM/MM methods have been used to investigate reactions of glutathione-S-transferase (GST) enzymes. GSTs are involved in the detoxification of many xenobiotic compounds, including many carcinogens and toxins in mammals. QM/MM molecular dynamics simulations of the conjugation of glutathione to phenanthrene 9,10 oxide in a glutathione-S-transferase, pinpointed a determinant of stereospecificity in this epoxide ring opening.[224] Similarly, specifically parameterized QM/MM methods have been applied to study the reaction between glutathione and 1-chloro-2,4-dinitrobenzene, which is a standard activity assay for glutathione S-transferases. QM/MM umbrella sampling molecular dynamics simulations of this reaction in the M1-1 GST isoenzyme, in mutant enzymes, and in solution, gave results in very good agreement with experiments.[93]

The human scavenger decapping enzyme (DcpS) is a dimeric enzyme involved in mRNA turnover. Crystal structures suggest that DcpS must undergo conformational changes upon ligand binding.[225,226] Long time-scale (20 ns) MD simulations of the apo-form of DcpS reveal a cooperative motion, with one active site closing whilst the other one opens.[227] This is an important example of a conformational change relevant to reaction being identified with MD.

The structures and other data generated by modelling enzyme-catalyzed reactions is potentially useful, *e.g.* in structure-based drug design.[20,228] An important future goal will be to make simulation methods and results accessible to non-specialists. Molecular calculations will contribute significantly to efforts to develop databases of enzyme mechanisms,[229] both in testing literature mechanisms and in adding structural and other (*e.g.* electronic properties) dimensions, in particular for reaction intermediates and transition states.

## 5. Conclusions

Molecular modelling and simulation are excellent ways to investigate the mechanisms of enzyme-catalysed reactions, and enzyme specificity and catalysis. Powerful techniques are now available that can provide uniquely detailed analysis, and useful practical insight, into the fundamental processes involved in biological catalysts. Simulations complement experiments, and will increasingly be important in the study of enzymes. Modelling enzymes presents many challenges, making careful testing and validation important. One exciting new area in which molecular modelling will play a vital part is catalyst design.[230–234] The design of protein or biomimetic catalysts with specific activities is increasingly a realistic goal, and reliable modelling methods have a vital part to play in these efforts.

## Acknowledgements

AJM is an EPSRC Leadership Fellow, and (with RL and KER) thanks EPSRC for support. AJM thanks his co-workers in the work described here. Some of this work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol (http://www.acrc.bris.ac.uk/acrc/index.htm).

## Notes and references

1 D. J. Vocadlo, G. J. Davies, R. Laine and S. G. Withers, *Nature*, 2001, **412**, 835–838.
2 A. L. Bowman, I. M. Grant and A. J. Mulholland, *Chem. Commun.*, 2008, 4425–4427.
3 W. W. Cleland, P. A. Frey and J. A. Gerlt, *J. Biol. Chem.*, 1998, **273**, 25529–25532.
4 A. J. Mulholland, P. D. Lyne and M. Karplus, *J. Am. Chem. Soc.*, 2000, **122**, 534–535.
5 C. N. Schutz and A. Warshel, *Proteins: Struct., Funct., Bioinf.*, 2004, **55**, 711–723.
6 P. A. Molina and J. H. Jensen, *J. Phys. Chem. B*, 2003, **107**, 6226–6233.
7 S. Hur and T. C. Bruice, *J. Am. Chem. Soc.*, 2003, **125**, 10540–10542.
8 M. Strajbl, A. Shurki, M. Kato and A. Warshel, *J. Am. Chem. Soc.*, 2003, **125**, 10228–10237.
9 K. E. Ranaghan and A. J. Mulholland, *Chem. Commun.*, 2004, 1238–1239.
10 M. H. M. Olsson, W. W. Parson and A. Warshel, *Chem. Rev.*, 2006, **106**, 1737–1756.
11 S. Hammes-Schiffer and J. B. Watney, *Philos. Trans. R. Soc. London, Ser. B*, 2006, **361**, 1365–1373.
12 L. Masgrau, A. Roujeinikova, L. O. Johannissen, P. Hothi, J. Basran, K. E. Ranaghan, A. J. Mulholland, M. J. Sutcliffe, N. S. Scrutton and D. Leys, *Science*, 2006, **312**, 237–241.
13 E. Hatcher, A. V. Soudackov and S. Hammes-Schiffer, *J. Am. Chem. Soc.*, 2007, **129**, 187–196.
14 H. H. Limbach, J. M. Lopez and A. Kohen, *Philos. Trans. R. Soc. London, Ser. B*, 2006, **361**, 1399–1415.
15 Z. D. Nagel and J. P. Klinman, *Chem. Rev.*, 2006, **106**, 3095–3118.
16 K. E. Ranaghan and A. J. Mulholland, *Interdiscip. Sci. Comput. Life Sci.*, 2010, **2**, 78–97.
17 A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu and M. H. M. Olsson, *Chem. Rev.*, 2006, **106**, 3210–3235.
18 J. Villa, M. Strajbl, T. M. Glennon, Y. Y. Sham, Z. T. Chu and A. Warshel, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 11899–11904.
19 A. J. Mulholland, in *Chemical Modelling and Theory*, ed. A. Hinchliffe, RSC Specialist Periodical Reports, 2006, pp. 23–68.
20 A. J. Mulholland, *Drug Discovery Today*, 2005, **10**, 1393–1402.
21 M. Garcia-Viloca, J. Gao, M. Karplus and D. G. Truhlar, *Science*, 2004, **303**, 186–195.
22 A. Warshel, *Annu. Rev. Biophys. Biomol. Struct.*, 2003, **32**, 425–443.
23 R. A. Friesner and V. Guallar, *Annu. Rev. Phys. Chem.*, 2005, **56**, 389–427.
24 T. C. Bruice and K. Kahn, *Curr. Opin. Chem. Biol.*, 2000, **4**, 540–544.
25 S. Martí, M. Roca, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *Chem. Soc. Rev.*, 2004, **33**, 98–107.
26 F. Himo, *Theor. Chem. Acc.*, 2006, **116**, 232–240.
27 H. M. Senn and W. Thiel, *Top. Curr. Chem.*, 2007, **268**, 173–290.
28 P. E. M. Siegbahn and T. Borowski, *Acc. Chem. Res.*, 2006, **39**, 729–738.
29 M. W. van der Kamp, K. E. Shaw, C. J. Woods and A. J. Mulholland, *J. R. Soc. Interface*, 2008, **5**, S173–190.
30 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227–249.
31 S. Scheiner and W. N. Lipscomb, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 432–436.
32 A. J. Mulholland, G. H. Grant and W. G. Richards, *Protein Eng.*, 1993, **6**, 133–147.
33 A. J. Mulholland and M. Karplus, *Biochem. Soc. Trans.*, 1996, **24**, 247–254.
34 J. Åqvist and A. Warshel, *Chem. Rev.*, 1993, **93**, 2523–2544.
35 H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198–1229.
36 S. C. L. Kamerlin, M. Haranczyk and A. Warshel, *J. Phys. Chem. B*, 2009, **113**, 1253–1272.
37 H. Lin and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **117**, 185–199.
38 F. Claeyssens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schütz, S. Thiel,

W. Thiel and H.-J. Werner, *Angew. Chem., Int. Ed.*, 2006, **45**, 6856–6859.
39 A. J. Mulholland, *Chem. Cent. J.*, 2007, **1**, 19.
40 S. Braun-Sand, M. H. M. Olsson and A. Warshel, *Adv. Phys. Org. Chem.*, 2005, **40**, 201–245.
41 S. Bjelic and J. Åqvist, *Biochemistry*, 2004, **43**, 14521–14528.
42 L. Ridder, J. N. Harvey, I. M. C. M. Rietjens, J. Vervoort and A. J. Mulholland, *J. Phys. Chem. B*, 2003, **107**, 2118–2126.
43 L. Ridder, A. J. Mulholland, I. M. C. M. Rietjens and J. Vervoort, *J. Am. Chem. Soc.*, 2000, **122**, 8728–8738.
44 M. Karplus, Y. Q. Gao, J. P. Ma, A. van der Vaart and W. Yang, *Philos. Trans. R. Soc. London, Ser. A*, 2005, **363**, 331–355.
45 A. Fersht, *Structure and Mechanism in Protein Science. A Guide to Enzyme Catalysis and Protein Folding*, Freeman, New York, 1999.
46 M. H. M. Olsson and A. Warshel, *J. Am. Chem. Soc.*, 2004, **126**, 15167–15179.
47 M. Wolf-Watz, V. Thai, K. Henzler-Wildman, G. Hadjipaylou, E. Z. Eisenmesser and D. Kern, *Nat. Struct. Mol. Biol.*, 2004, **11**, 945–949.
48 A. Kohen, R. Cannio, S. Bartolucci and J. P. Klinman, *Nature*, 1999, **399**, 496–499.
49 M. J. Sutcliffe, L. Masgrau, A. Roujeinikova, L. O. Johannissen, P. Hothi, J. Basran, K. E. Ranaghan, A. J. Mulholland, D. Leys and N. S. Scrutton, *Philos. Trans. R. Soc. London, Ser. B*, 2006, **361**, 1375–1386.
50 J. Bentzien, R. P. Muller, J. Florian and A. Warshel, *J. Phys. Chem. B*, 1998, **102**, 2293–2301.
51 S. C. L. Kamerlin, J. Cao, E. Rosta and A. Warshel, *J. Phys. Chem. B*, 2009, **113**, 10905–10915.
52 G. Y. Hong, E. Rosta and A. Warshel, *J. Phys. Chem. B*, 2006, **110**, 19570–19574.
53 A. Warshel, *Computer modeling of Chemical Reactions in Enzymes and Solutions*, John Wiley & Sons, New York, 1997.
54 J. Villà and A. Warshel, *J. Phys. Chem. B*, 2001, **105**, 7887–7907.
55 J. Florian, M. F. Goodman and A. Warshel, *J. Am. Chem. Soc.*, 2003, **125**, 8163–8177.
56 P. Varnai and A. Warshel, *J. Am. Chem. Soc.*, 2000, **122**, 3849–3860.
57 A. Warshel, P. K. Sharma, Z. T. Chu and J. Åqvist, *Biochemistry*, 2007, **46**, 1466–1476.
58 S. Bjelic and J. Åqvist, *Biochemistry*, 2006, **45**, 7709–7723.
59 S. Trobro and J. Åqvist, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 12395–12400.
60 P. K. Sharma, Y. Xiang, M. Kato and A. Warshel, *Biochemistry*, 2005, **44**, 11307–11314.
61 S. Hammes-Schiffer, *Curr. Opin. Struct. Biol.*, 2004, **14**, 192–201.
62 H. B. Liu and A. Warshel, *J. Phys. Chem. B*, 2007, **111**, 7852–7861.
63 P. K. Sharma, Z. T. Chu, M. H. M. Olsson and A. Warshel, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 9661–9666.
64 M. Roca, A. Vardi-Kilshtain and A. Warshel, *Biochemistry*, 2009, **48**, 3046–3056.
65 M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902–3909.
66 J. J. P. Stewart, *J. Comput. Chem.*, 1989, **10**, 209–220.
67 A. van der Vaart, V. Gogonea, S. L. Dixon and K. M. Merz Jr., *J. Comput. Chem.*, 2000, **21**, 1494–1504.
68 J. Khandogin and D. M. York, *Proteins: Struct., Funct., Bioinf.*, 2004, **56**, 724–737.
69 J. Khandogin, K. Musier-Forsyth and D. M. York, *J. Mol. Biol.*, 2003, **330**, 993–1004.
70 J. C. Hermann, C. Hensen, L. Ridder, A. J. Mulholland and H.-D. Höltje, *J. Am. Chem. Soc.*, 2005, **127**, 4454–4465.
71 F. Jensen, *in Introduction to Computational Chemistry*, Wiley, Chichester, 2006, pp. 115–131.
72 S. Grimme, J. Antony, T. Schwabe and C. Mueck-Lichtenfeld, *Org. Biomol. Chem.*, 2007, **5**, 741–758.
73 F. Himo and P. E. M. Siegbahn, *Chem. Rev.*, 2003, **103**, 2421–2456.
74 P. E. M. Siegbahn and F. Himo, *J. Biol. Inorg. Chem.*, 2009, **14**, 643–651.
75 A. Shurki and A. Warshel, *Adv. Protein Chem.*, 2003, **66**, 249.
76 Y. Zhang, J. Kua and J. A. McCammon, *J. Phys. Chem. B*, 2003, **107**, 4459–4463.
77 A. Lodola, M. Mor, J. Żurek, G. Tarzia, D. Piomelli, J. N. Harvey and A. J. Mulholland, *Biophys. J.*, 2007, **92**, L20–L22.
78 J. L. Gao and D. G. Truhlar, *Annu. Rev. Phys. Chem.*, 2002, **53**, 467–505.
79 C. J. Woods, F. R. Manby and A. J. Mulholland, *J. Chem. Phys.*, 2008, **128**, 014109.
80 M. Garcia-Viloca, T. D. Poulsen, D. G. Truhlar and J. Gao, *Protein Sci.*, 2004, **13**, 2341–2354.
81 K. E. Ranaghan, L. Ridder, B. Szefczyk, W. A. Sokalski, J. C. Hermann and A. J. Mulholland, *Mol. Phys.*, 2003, **101**, 2695–2714.
82 H. Lee Woodcock, M. Hodoscek, P. Sherwood, Y. S. Lee, H. F. Schaefer and B. R. Brooks, *Theor. Chem. Acc.*, 2003, **109**, 140–148.
83 M. J. Field, P. A. Bash and M. Karplus, *J. Comput. Chem.*, 1990, **11**, 700–733.
84 P. D. Lyne, M. Hodoscek and M. Karplus, *J. Phys. Chem. A*, 1999, **103**, 3462–3471.
85 Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim and M. Karplus, *J. Phys. Chem. B*, 2001, **105**, 569–585.
86 M. J. Field, *J. Comput. Chem.*, 2002, **23**, 48.
87 K. E. Ranaghan and A. J. Mulholland, *Int. Rev. Phys. Chem.*, 2010, **29**, 65–133.
88 K. E. Ranaghan, L. Ridder, B. Szefczyk, W. A. Sokalski, J. C. Hermann and A. J. Mulholland, *Org. Biomol. Chem.*, 2004, **2**, 968–980.
89 D. Riccardi, P. Schaefer and Q. Cui, *J. Phys. Chem. B*, 2005, **109**, 17715.
90 K. Raha and K. M. Merz Jr, *J. Am. Chem. Soc.*, 2004, **126**, 1020.
91 S. Marti and V. Moliner, *J. Chem. Theory Comput.*, 2005, **1**, 1008–1016.
92 X. Prat-Resina, J. M. Bofill, A. Gonzalez-Lafont and J. M. Lluch, *Int. J. Quantum Chem.*, 2004, **98**, 367–377.
93 A. L. Bowman, L. Ridder, I. M. C. M. Rietjens, J. Vervoort and A. J. Mulholland, *Biochemistry*, 2007, **46**, 6353–6363.
94 J. Gao, S. Ma, D. T. Major, K. Nam, J. Pu and D. G. Truhlar, *Chem. Rev.*, 2006, **106**, 3188–3209.
95 M. Klahn, S. Braun-Sand, E. Rosta and A. Warshel, *J. Phys. Chem. B*, 2005, **109**, 15645–15650.
96 D. Bakowies and W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580.
97 M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber and K. Morokuma, *J. Phys. Chem.*, 1996, **100**, 19357.
98 W. Yang and D. G. Drueckhammer, *J. Phys. Chem. B*, 2003, **107**, 5986.
99 K. Senthilkumar, J. I. Mujika, K. E. Ranaghan, F. R. Manby, A. J. Mulholland and J. N. Harvey, *J. R. Soc. Interface*, 2008, **5**, S207–S216.
100 I. Antes and W. Thiel, *J. Phys. Chem. A*, 1999, **103**, 9290.
101 L. Jensen and P. T. van Duijnen, *J. Chem. Phys.*, 2005, **123**, 074307.
102 S. P. Greatbanks, J. E. Gready, A. C. Limaye and A. P. Rendell, *Proteins: Struct., Funct., Genet.*, 1999, **37**, 157.
103 L. Ridder and A. J. Mulholland, *Curr. Top. Med. Chem.*, 2003, **3**, 1241–1256.
104 C. B. Post and M. Karplus, *J. Am. Chem. Soc.*, 1986, **108**, 1317–1319.
105 K. E. Ranaghan, L. Masgrau, N. S. Scrutton, M. J. Sutcliffe and A. J. Mulholland, *ChemPhysChem*, 2007, **8**, 1816–1835.
106 P. Kast, M. Asif-Ullah and D. Hilvert, *Tetrahedron Lett.*, 1996, **37**, 2691–2694.
107 P. D. Lyne, A. J. Mulholland and W. G. Richards, *J. Am. Chem. Soc.*, 1995, **117**, 11345–11350.
108 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *J. Phys. Chem. B*, 2000, **104**, 11308.
109 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *Theor. Chem. Acc.*, 2001, **105**, 207–212.
110 B. Szefczyk, A. J. Mulholland, K. E. Ranaghan and W. A. Sokalski, *J. Am. Chem. Soc.*, 2004, **126**, 16148–16159.
111 C. R. W. Guimaraes, M. Udier-Blagovic, I. Tubert-Brohman and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2005, **1**, 617–625.
112 A. Kienhöfer, P. Kast and D. Hilvert, *J. Am. Chem. Soc.*, 2003, **125**, 3206–3207.
113 H. Guo, Q. Cui, W. N. Lipscomb and M. Karplus, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 9032.
114 S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón and J. Bertrán, *Chem. Eur. J.*, 2003, **9**, 984.
115 C. R. W. Guimarães, M. P. Repasky, J. Chandrasekhar, J. Tirado-Rives and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2003, **125**, 6892–6899.

116 S. Hur and T. C. Bruice, *J. Am. Chem. Soc.*, 2003, **125**, 5964–5972.
117 S. Hur and T. C. Bruice, *J. Am. Chem. Soc.*, 2003, **125**, 1472–1473.
118 J. B. S. Haldane, in *Enzymes*, Longmans Green and Co., 1930, p. 182.
119 F. Claeyssens, K. E. Ranaghan, F. R. Manby, J. N. Harvey and A. J. Mulholland, *Chem. Commun.*, 2005, 5068–5070.
120 J. N. Harvey, *Faraday Discuss.*, 2004, **127**, 165–177.
121 *Jaguar 5.5*, Schrödinger LLC, Portland, OR, 2001, http://www.schrodinger.com.
122 J. W. Ponder, Tinker - Software Tools for Molecular Design, 2004, http://dasher.wustl.edu/tinker/.
123 A. Crespo, D. A. Scherlis, M. A. Marti, P. Ordejon, A. E. Roitberg and D. A. Estrin, *J. Phys. Chem. B*, 2003, **107**, 13728–13736.
124 Y. S. Lee, S. E. Worthington, M. Krauss and B. R. Brooks, *J. Phys. Chem. B*, 2002, **106**, 12059–12065.
125 F. Manby, H.-J. Werner, T. Adler and A. May, *J. Chem. Phys.*, 2006, **124**, 094103.
126 H.-J. Werner, F. Manby and P. Knowles, *J. Chem. Phys.*, 2003, **118**, 8149.
127 D. C. Phillips, *Proc. Nat. Acad. Sci. U.S.A*, 1967, **57**, 484–495.
128 D. E. Koshland, *Biol. Rev.*, 1953, **28**, 416–436.
129 A. J. Kirby, *Nat. Struct. Biol.*, 2001, **8**, 737–739.
130 A. Bottoni, G. P. Miscione and M. D. Vivo, *Proteins: Struct., Funct., Bioinf.*, 2005, **59**, 118–130.
131 A. D. MacKerell, D. Bashford, R. L. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
132 D. M. Chipman, *Biochemistry*, 1971, **10**, 1714–1722.
133 S. J. Remington, *Curr. Top. Cell. Regul.*, 1992, **33**, 209.
134 M. Karpusas, B. Branchaud and S. J. Remington, *Biochemistry*, 1990, **29**, 2213.
135 G. Löhlein-Werhahn, E. Bayer, B. Bauer and H. Eggerer, *Eur. J. Biochem.*, 1983, **133**, 665.
136 A. J. Mulholland and W. G. Richards, *Proteins: Struct., Funct., Genet.*, 1997, **27**, 9.
137 A. J. Mulholland and W. G. Richards, *THEOCHEM*, 1998, **427**, 175.
138 M. W. van der Kamp, F. Perruccio and A. J. Mulholland, *J. Mol. Graphics Modell.*, 2007, **26**, 596.
139 S. Bjelic, B. O. Brandsdal and J. Åqvist, *Biochemistry*, 2008, **47**, 10049–10057.
140 O. Donini, T. Darden and P. Kollman, *J. Am. Chem. Soc.*, 2000, **122**, 12270–12280.
141 A. J. Mulholland and W. G. Richards, *J. Phys. Chem. B*, 1998, **102**, 6635–6646.
142 M. W. van der Kamp, F. Perruccio and A. J. Mulholland, *Chem. Commun.*, 2008, 1874–1876.
143 G. M. Alter, J. P. Casazza, Z. Wang, P. Nemeth, P. A. Srere and C. T. Evans, *Biochemistry*, 1990, **29**, 7557.
144 A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman and Co., New York, 1999.
145 L. Kurz and G. Drysdale, *Biochemistry*, 1987, **26**, 2623–2627.
146 L. Kurz, J. Ackerman and G. Drysdale, *Biochemistry*, 1985, **24**, 452–457.
147 J. N. Harvey, C. M. Bathelt and A. J. Mulholland, *J. Comput. Chem.*, 2006, **27**, 1352–1362.
148 B. Meunier, S. P. de Visser and S. Shaik, *Chem. Rev.*, 2004, **104**, 3947–3980.
149 S. Shaik, D. Kumar, S. P. de Visser, A. Altun and W. Thiel, *Chem. Rev.*, 2005, **105**, 2279–2328.
150 F. Ogliaro, N. Harris, S. Cohen, M. Filatov, S. P. de Visser and S. Shaik, *J. Am. Chem. Soc.*, 2000, **122**, 8977–8989.
151 N. Harris, S. Cohen, M. Filatov, F. Ogliaro and S. Shaik, *Angew. Chem., Int. Ed.*, 2000, **39**, 2003.
152 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle and J. A. Pople, GAUSSIAN, Inc., Pittsburgh PA, see www.gaussian.com.
153 K. Yoshizawa, T. Kamachi and Y. Shiota, *J. Am. Chem. Soc.*, 2001, **123**, 9806–9816.
154 S. P. de Visser, F. Ogliaro, N. Harris and S. Shaik, *J. Am. Chem. Soc.*, 2001, **123**, 3037–3047.
155 S. P. de Visser, F. Ogliaro and S. Shaik, *Chem. Commun.*, 2001, 2322–2323.
156 F. P. Guengerich, *Chem. Res. Toxicol.*, 2001, **14**, 611–650.
157 C. M. Bathelt, L. Ridder, A. J. Mulholland and J. N. Harvey, *J. Am. Chem. Soc.*, 2003, **125**, 15004–15005.
158 C. M. Bathelt, L. Ridder, A. J. Mulholland and J. N. Harvey, *Org. Biomol. Chem.*, 2004, **2**, 2998–3005.
159 M. J. de Groot, S. B. Kirton and M. J. Sutcliffe, *Curr. Top. Med. Chem.*, 2004, **4**, 1803–1824.
160 J. N. Harvey, V. K. Aggarwal, C. M. Bathelt, J.-L. Carreón-Macedo, T. Gallagher, N. Holzmann, A. J. Mulholland and R. Robiette, *J. Phys. Org. Chem.*, 2006, **19**, 608–615.
161 M. Pirmohamed and B. K. Park, *Toxicology*, 2003, **192**, 23–32.
162 C. M. Bathelt, J. Zurek, A. J. Mulholland and J. N. Harvey, *J. Am. Chem. Soc.*, 2005, **127**, 12900–12908.
163 C. M. Bathelt, A. J. Mulholland and J. N. Harvey, *J. Phys. Chem. A*, 2008, **112**, 13149–13156.
164 V. Nedelcheva, I. Gut, P. Souček, B. Tichavská, L. Týnkova, J. Mráz, F. P. Guengerich and M. Ingelman-Sundberg, *Arch. Toxicol.*, 1999, **73**, 33–40.
165 J. C. Schoneboom, S. Cohen, H. Lin, S. Shaik and W. Thiel, *J. Am. Chem. Soc.*, 2004, **126**, 4017–4034.
166 V. Guallar, M.-H. Baik, S. J. Lippard and R. A. Friesner, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 6998–7002.
167 J. C. Schoneboom, H. Lin, N. Reuter, W. Thiel, S. Cohen, F. Ogliaro and S. Shaik, *J. Am. Chem. Soc.*, 2002, **124**, 8142–8151.
168 J. C. Schoneboom, F. Neese and W. Thiel, *J. Am. Chem. Soc.*, 2005, **127**, 5840–5853.
169 J. C. Schoneboom and W. Thiel, *J. Phys. Chem. B*, 2004, **108**, 7468–7478.
170 H. Lin, J. C. Schoneboom, S. Cohen, S. Shaik and W. Thiel, *J. Phys. Chem. B*, 2004, **108**, 10083–10088.
171 V. Guallar and R. A. Friesner, *J. Am. Chem. Soc.*, 2004, **126**, 8501–8508.
172 M. Altarsha, T. Benighaus, D. Kumar and W. Thiel, *J. Am. Chem. Soc.*, 2009, **131**, 4755–4763.
173 A. Altun, V. Guallar, R. A. Friesner, S. Shaik and W. Thiel, *J. Am. Chem. Soc.*, 2006, **128**, 3924–3925.
174 V. Guallar and B. Olsen, *J. Inorg. Biochem.*, 2006, **100**, 755–760.
175 A. Altun, S. Shaik and W. Thiel, *J. Comput. Chem.*, 2006, **27**, 1324–1337.
176 J. Żurek, N. Foloppe, J. N. Harvey and A. J. Mulholland, *Org. Biomol. Chem.*, 2006, **4**, 3931–3937.
177 R. Lonsdale, J. N. Harvey and A. J. Mulholland, *J. Phys. Chem. B*, 2010, **114**, 1156–1162.
178 S. C. Tsai and J. P. Klinman, *Biochemistry*, 2001, **40**, 2303.
179 J. Basran, S. Patel, M. J. Sutcliffe and N. S. Scrutton, *J. Biol. Chem.*, 2001, **276**, 6234.
180 P. F. Faulder, G. Tresadern, K. K. Chohan, N. S. Scrutton, M. J. Sutcliffe, I. H. Hillier and N. A. Burton, *J. Am. Chem. Soc.*, 2001, **123**, 8604–8605.
181 G. Tresadern, J. P. McNamara, M. Mohr, H. Wang, N. A. Burton and I. H. Hillier, *Chem. Phys. Lett.*, 2002, **358**, 489–494.
182 G. Tresadern, S. Nunez, P. F. Faulder, H. Wang, I. H. Hillier and N. A. Burton, *Faraday Discuss.*, 2003, **122**, 223–242.
183 G. Pierdominici-Sottile, J. Echave and J. Palma, *Int. J. Quantum Chem.*, 2005, **105**, 937–945.

184 G. Pierdominici-Sottile, J. Echave and J. Palma, *J. Phys. Chem. B*, 2006, **110**, 11592–11599.
185 C. Alhambra, M. L. Sanchez, J. C. Corchado, J. Gao and D. G. Truhlar, *Chem. Phys. Lett.*, 2002, **355**, 388–394.
186 J. Z. Pu, J. L. Gao and D. G. Truhlar, *J. Phys. Chem. A*, 2004, **108**, 632–650.
187 M. Garcia-Viloca, C. Alhambra, D. G. Truhlar and J. Gao, *J. Chem. Phys.*, 2001, **114**, 9953.
188 C. Alhambra, J. Corchado, M. L. Sanchez, M. Garcia-Viloca, J. Gao and D. G. Truhlar, *J. Phys. Chem. B*, 2001, **105**, 11326.
189 J. K. Hwang, Z. T. Chu, A. Yadav and A. Warshel, *J. Phys. Chem.*, 1991, **95**, 8445–8448.
190 G. Tresadern, H. Wang, P. F. Faulder, N. A. Burton and I. H. Hillier, *Mol. Phys.*, 2003, **101**, 2775–2784.
191 J. Mavri, H. Liu, M. H. M. Olsson and A. Warshel, *J. Phys. Chem. B*, 2008, **112**, 5950–5954.
192 G. Pierdominici-Sottile and J. Palma, *Chem. Phys.*, 2009, **363**, 59–64.
193 H. B. Brooks, L. H. Jones and V. L. Davidson, *Biochemistry*, 1993, **32**, 2725–2729.
194 L. Masgrau, K. E. Ranaghan, N. S. Scrutton, A. J. Mulholland and M. J. Sutcliffe, *J. Phys. Chem. B*, 2007, **111**, 3032–3047.
195 A. Lodola, M. Mor, J. C. Hermann, G. Tarzia, D. Piomelli and A. J. Mulholland, *Chem. Commun.*, 2005, 4399–4401.
196 I. Tubert-Brohman, O. Acevedo and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2006, **128**, 16904–16913.
197 A. Lodola, M. Mor, S. Rivara, C. Christov, G. Tarzia, D. Piomelli and A. J. Mulholland, *Chem. Commun.*, 2008, 214–216.
198 D. Piomelli, G. Tarzia, A. Duranti, A. Tontini, M. Mor, T. R. Compton, O. Dasse, E. P. Monaghan, J. A. Parrott and D. Putman, *CNS Drug Rev.*, 2006, **12**, 21–38.
199 A. Bassan, M. R. A. Blomberg and P. E. M. Siegbahn, *JBIC, J. Biol. Inorg. Chem.*, 2004, **9**, 439–452.
200 K.-B. Cho, V. Pelmenschikov, A. Gräslund and P. E. M. Siegbahn, *J. Phys. Chem. B*, 2004, **108**, 2056–2065.
201 T. Borowski, A. Bassan and P. E. M. Siegbahn, *Biochemistry*, 2004, **43**, 12331–12342.
202 D. Xu, Y. Wei, J. Wu, D. Dunaway-Mariano, H. Guo, Q. Cui and J. Gao, *J. Am. Chem. Soc.*, 2004, **126**, 13649–13658.
203 G. H. Li and Q. Cui, *J. Am. Chem. Soc.*, 2003, **125**, 15028–15038.
204 S. Martí, V. Moliner, I. Tuñón and I. H. Williams, *J. Phys. Chem. B*, 2005, **109**, 3707–3710.
205 G. D. Ruggiero, I. H. Williams, M. Roca, V. Moliner and I. Tuñón, *J. Am. Chem. Soc.*, 2004, **126**, 8634–8635.
206 H. Park, E. N. Brothers and K. M. Merz Jr., *J. Am. Chem. Soc.*, 2005, **127**, 4232–4241.
207 J. C. Hermann, J. Pradon, J. N. Harvey and A. J. Mulholland, *J. Phys. Chem. A*, 2009, **113**, 11984–11994.
208 G. A. Cisneros, M. Wang, P. Silinski, M. C. Fitzgerald and W. Yang, *Biochemistry*, 2004, **43**, 6885–6892.
209 D. Riccardi, P. König, H. Guo and Q. Cui, *Biochemistry*, 2008, **47**, 2369–2378.
210 S. Metz, D. Wang and W. Thiel, *J. Am. Chem. Soc.*, 2009, **131**, 4628–4640.
211 M. W. van der Kamp and A. J. Mulholland, *Nat. Prod. Rep.*, 2008, **25**, 1001–1014.
212 A. Crespo, M. A. Martí, D. A. Estrin and A. E. Roitberg, *J. Am. Chem. Soc.*, 2005, **127**, 6940–6941.
213 M. E. S. Soliman, G. D. Ruggiero, J. J. Ruiz Pernía, I. R. Greig and I. H. Williams, *Org. Biomol. Chem.*, 2009, **7**, 460–468.
214 C. R. W. Guimarães, M. Udier-Blagovi and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2005, **127**, 3577–3588.
215 M. P. Gleeson, I. H. Hillier and N. A. Burton, *Org. Biomol. Chem.*, 2004, **2**, 2275–2280.
216 M. Mladenovic, K. Ansorg, R. F. Fink, W. Thiel, T. Schirmeister and B. Engels, *J. Phys. Chem. B*, 2008, **112**, 11798–11808.
217 A. R. Dinner, G. M. Blackburn and M. Karplus, *Nature*, 2001, **413**, 752–755.
218 M. H. M. Olsson, P. K. Sharma and A. Warshel, *FEBS Lett.*, 2005, **579**, 2026–2034.
219 I. Tejero, M. Garcia-Viloca, A. Gonzalez-Lafont, J. M. Lluch and D. M. York, *J. Phys. Chem. B*, 2006, **110**, 24708–24719.
220 M. H. M. Olsson, P. E. M. Siegbahn and A. Warshel, *J. Am. Chem. Soc.*, 2004, **126**, 2820–2828.
221 M. H. M. Olsson, J. Mavri and A. Warshel, *Philos. Trans. R. Soc. London, Ser. B*, 2006, **361**, 1417–1432.
222 C. M. Bathelt, A. J. Mulholland and J. N. Harvey, *Dalton Trans.*, 2005, 3470–3476.
223 S. P. de Visser, *Biochem. Soc. Trans.*, 2009, **37**, 373–377.
224 L. Ridder, I. M. C. M. Rietjens, J. Vervoort and A. J. Mulholland, *J. Am. Chem. Soc.*, 2002, **124**, 9926–9936.
225 N. Chen, M. A. Walsh, Y. Y. Liu, R. Parker and H. W. Song, *J. Mol. Biol.*, 2005, **347**, 707–718.
226 M. G. Gu, C. Fabrega, S. W. Liu, H. D. Liu, M. Kiledjian and C. D. Lima, *Mol. Cell*, 2004, **14**, 67–80.
227 U. Pentikäinen, O. T. Pentikäinen and A. J. Mulholland, *Proteins: Struct., Funct., Bioinf.*, 2008, **70**, 498–508.
228 M. P. Gleeson and D. Gleeson, *J. Chem. Inf. Model.*, 2009, **49**, 670–677.
229 G. L. Holliday, D. E. Almonacid, J. B. O. Mitchell and J. M. Thornton, *J. Mol. Biol.*, 2007, **372**, 1261–1277.
230 L. Jiang, E. Althoff, F. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. Gallaher, J. Betker, F. Tanaka, B. Iii, F. Carlos, D. Hilvert, K. Houk, B. Stoddard and D. Baker, *Science*, 2008, **319**, 1387.
231 S. Park and J. G. Saven, *Annu. Rep. Comput. Chem.*, 2005, **1**, 245.
232 B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard and D. Baker, *Science*, 2003, **302**, 1364–1368.
233 D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik and D. Baker, *Nature*, 2008, **453**, 190–U194.
234 A. N. Alexandrova, D. Röthlisberger, D. Baker and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2008, **130**, 15907–15915.