CHAPTER SIX

QM/MM Calculations on Proteins

U. Ryde[1]
Chemical Centre, Lund University, Lund, Sweden
[1]Corresponding author: e-mail address: ulf.ryde@teokem.lu.se

Contents

Abstract

In this chapter, I discuss combined quantum mechanics (QM) and molecular mechanics (MM; QM/MM) calculations for proteins. In QM/MM, a small but interesting part of the protein is treated by accurate QM methods, whereas the remainder is treated by faster MM methods. The prime problems with QM/MM calculations are bonds between the QM and MM systems, the selection of the QM system, and the local-minima problem. The two first problems can be solved by the big-QM approach, including in the QM calculation all groups within 4.5–6 Å of the active site and all buried charges in the protein. The third problem can be solved by calculating free energies. It is important to study QM/MM energy components to ensure that the results are stable and reliable. They can also be used to understand the reaction and the effect of the surroundings, eg, by dividing the catalytic effect into bonded, van der Waals, electrostatic, and geometric components and to deduce which parts of the protein contribute most to the catalysis. It should be ensured that the QM calculations are reliable and converged by extending the basis set to quadruple-zeta quality, including a proper treatment of dispersion, as well as relativistic, zero-point, thermal, and entropy effects, and performing

Methods in Enzymology, Volume 577
ISSN 0076-6879
http://dx.doi.org/10.1016/bs.mie.2016.05.014

calculations with both pure and hybrid density functional theory methods. If the latter give differing results, calibration with high-level QM methods is needed. Reactions that change the net charge should be avoided. QM/MM calculations can be combined with experimental methods.

1. INTRODUCTION

During the latest two decades, quantum mechanical (QM) calculations have been established as a powerful complement to experiments for the study of structural and energetic properties of proteins and enzyme reactions. QM calculations have the advantage of directly providing energetic information, which govern all chemical processes. Moreover, they are not subject to experimental limitations, such as invisible, short-lived, expensive, or hazardous compounds. In particular, QM calculations have been used to characterize and study the transition states and activation energies of enzyme reaction, allowing for the discrimination between alternative suggested reaction mechanisms. With QM calculations, reactions are studied by computers, avoiding all use of chemicals, which make them the ultimate type of environment-friendly green chemistry. On the other hand, QM calculations are approximate, so the results should be calibrated against experimental data and you should ensure that the calculations are converged.

There are two approaches to treat proteins with QM methods. In the QM-cluster approach, a small number of residues (typically 30–200 atoms; cf. Fig. 1A) are cut out from the protein and are studied by QM in a continuum solvent (Blomberg, Borowski, Himo, Liao, & Siegbahn, 2014; Himo & Siegbahn, 2009). This allows for a full control of the structures of the studied system. On the other hand, the results may be biased by the choice of included residues and the surrounding protein is studied in a very crude manner. Alternatively, the active site is treated by a similar QM model, but the surrounding protein and solvent are modeled by molecular mechanics (MM), giving the QM/MM approach (Lin & Truhlar, 2007; Senn & Thiel, 2009; Sousa, Fernandes, & Ramos, 2012; Fig. 1B). This provides a more detailed description of the surroundings, but the calculations become more complicated, and involve thousands of atoms, which cannot be easily controlled.

In this chapter, I provide my view of how QM/MM studies of proteins should be performed to give useful and reliable results, based on my more than 20 years experience and method development (Delcey et al., 2014;

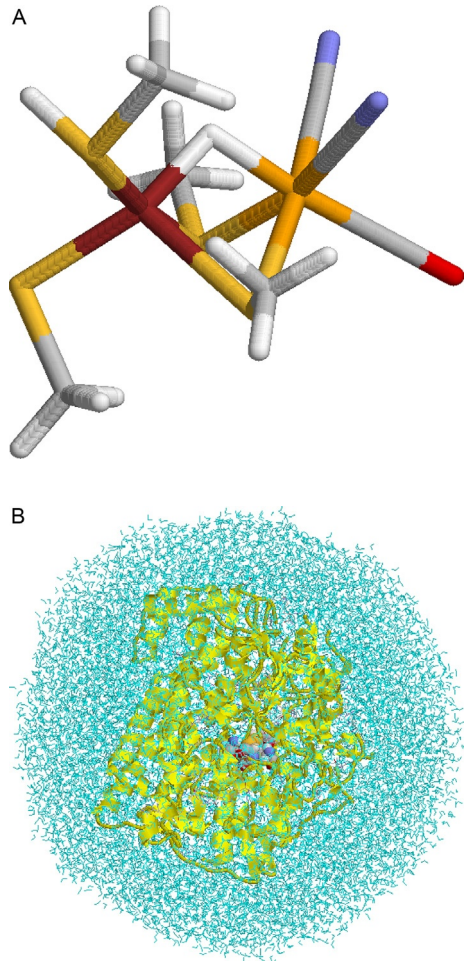

Fig. 1 Examples of (A) QM-cluster, (B) QM/MM (the QM system is shown as *balls*), and

Dong & Ryde, 2016; Hedegård, Kongsted, & Ryde, 2015; Heimdal, Kaukonen, Srnec, Rulíšek, & Ryde, 2011; Hu, Söderhjelm, & Ryde, 2011, 2013; Kaukonen, Söderhjelm, Heimdal, & Ryde, 2008a; Li, Farrokhnia, Rulíšek, & Ryde, 2015; Rod & Ryde, 2005a; Ryde, 1995, 1996; Ryde & Olsson, 2001; Ryde, Olsen, & Nilsson, 2002; Söderhjelm & Ryde, 2006; Sumner, Söderhjelm, & Ryde, 2013). I start with a short description of the methods involved and the special approaches

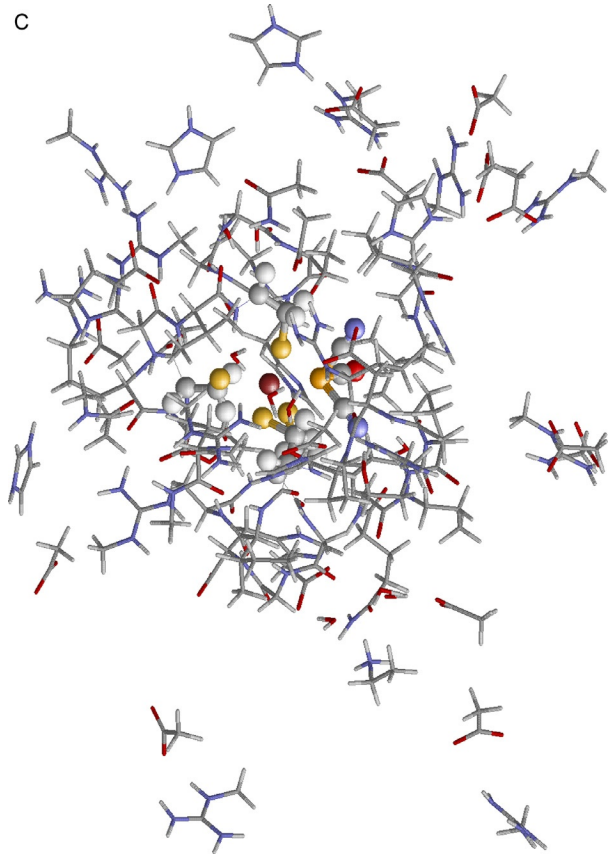

Fig. 1—Cont'd (C) big-QM models for [NiFe] hydrogenase (Dong & Ryde, 2016).

developed in our group. Then, I give some typical applications, illustrating various aspects of these methods. In the final section, I provide a detailed scheme of the recommended QM/MM approach, together with some concluding remarks.

2. METHODS

In this section, an introduction to the methods employed is given. For QM and MM methods, only the very basics are given and the interested reader is referred to textbooks in computational chemistry (eg, Cramer, 2006; Jensen, 2007). For the more specific QM/MM approaches, a

somewhat deeper introduction is given, but for technical details, the reader is referred to review articles (Lin & Truhlar, 2007; Senn & Thiel, 2009; Sousa et al., 2012).

2.1 QM Methods

In QM calculations, the Schrödinger equation is solved for a molecular system (Cramer, 2006; Jensen, 2007). The input is the coordinates, the net charge, and the total spin. The result is the total energy and the wavefunction, from which all measureable properties of the system can be calculated. Unfortunately, the Schrödinger equation can be analytically solved only for a few systems with a single electron—for all large systems, only approximate numerical solutions can be obtained. The most basic QM approach is the Hartree–Fock (HF) method (Fock, 1930; Hartree, 1928). Unfortunately, it gives quite approximate results and it is therefore mainly used as a starting point of more accurate methods, based on perturbation theory or series expansions, eg, Møller–Plesset second-order perturbation theory (MP2; Møller & Plesset, 1934) or coupled cluster calculations with single, double, and perturbatively treated triple excitations (CCSD(T); Raghavachari, Tucks, Pople, & Head-Gordon, 1989). The latter is currently considered as the gold standard QM method. Unfortunately, it is computationally a very demanding method so that it in practice can be used only for molecules with up to ~30 atoms. However, recently methods have been developed that take advantage of the local nature of the wavefunction, eg, LCCSD(T0), which can be used for systems with more than 100 atoms (Werner & Schütz, 2011).

Today, the great majority of QM calculations are performed with density functional theory (DFT). It is not based on the wavefunction, which is a function of the coordinates of all involved particles, but instead on the electron density, which is a function of only the three coordinates of Cartesian space. Still, there is a one-to-one correspondence between a wavefunction and the electron density (Hohenberg & Kohn, 1964). DFT also involves approximations and there are a great number of variants, eg, BP86 (Becke, 1988; Perdew, 1986), PBE (Perdew, Burke, & Ernzerhof, 1996), and TPSS (Tao, Perdew, Staroverov, & Scuseria, 2003). So-called hybrid DFT methods involve a fraction of exchange from the HF method, eg, B3LYP (Becke, 1993; Lee, Yang, & Parr, 1988). In general, DFT methods give results of MP2 quality or better at a lower computational effort.

All mentioned QM methods expand the wavefunction in a set of known functions, the basis set. In general, the larger the basis set, the more accurate will the results be. Reasonable structures can be obtained with one basis function for the core electrons, two for the valence electrons, and one set of basis function with an angular momentum one step higher than for the valence electrons for nonhydrogen atoms. This is called a polarized split-valence basis set (SVP), eg, def2-SV(P) (Weigend & Ahlrichs, 2005). Energies normally require at least three basis functions for the valence electrons (valence triple-zeta basis sets, VTZ), eg, def2-TZVP and the convergence should be checked with four basis functions for the valence electrons (valence quadruple-zeta basis sets), eg, def2-QZVP (Weigend & Ahlrichs, 2005). Anions require the use of more diffuse basis functions, eg, def2-QZVPD (Rappoport & Furche, 2010).

QM calculations can be sped up by a factor of ~1000 by using a minimal basis set and replacing all integrals by empirical parameters, the semiempirical QM methods (SEQM; Thiel, 2014). Many such methods have been suggested, eg, AM1 (Dewar, Zoebisch, Healy, & Stewart, 1985), PM3 (Stewart, 1989), and PM6 (Stewart, 2007), but the accuracy is typically lower than for QM and DFT methods and somewhat unpredictable, especially for unusual electronic structures, such as transition states.

Neither HF nor DFT calculations provide a proper description of the London dispersion interactions. Several approaches have been suggested to correct this problem (Grimme, 2011), but Grimme's DFT-D2 and D3 methods are most widely used (Grimme, 2006; Grimme, Antony, Ehrlich, & Krieg, 2010).

QM calculations are normally performed on isolated molecules in gas phase. However, most experiments are performed in condensed phases, eg, water solution. Methods have been developed to model a surrounding homogeneous solvent in QM calculations, continuum solvation methods (Tomasi, Mennucci, & Cammi, 2005), eg, the polarizable continuum model (PCM) (Cossi, Tomasi, & Cammi, 1995), or the conductor-like solvent model (COSMO) (Klamt & Schüürmann, 1994). For protein-sized systems, methods based on the Poisson–Boltzmann (PB) equation or the generalized Born (GB) approach are more common (Bashford & Case, 2000; Honig, Sharp, & Yang, 1997).

2.2 MM Methods

In MM methods, no attempt is made to solve the Schrödinger equation and electrons are ignored. Instead, molecules are considered as a collection of

balls, connected by springs. The energy of the system as a function of the coordinates is described by an empirical function, a force field. For proteins, it typically contains terms for the distortion of bonds, angles, and dihedrals, as well as the nonbonded exchange-repulsion, dispersion (van der Waals), and electrostatic interaction energies (Mackerell, 2004). Such an energy can be calculated for a whole protein in seconds, allowing for extensive sampling of the accessible phase space by molecular dynamics (MD) or Monte Carlo methods. The protein is typically solvated with several thousands of explicit water molecules and possibly counterions to provide a more realistic account of the surroundings. To mimic infinite systems, periodic boundary conditions are employed and long-range electrostatic interactions are often treated by Ewald summation (Darden, York, & Pedersen, 1993).

2.3 QM/MM Calculations

The philosophy behind the QM/MM approach is that a QM method is used for a small, but interesting, part of the protein, eg, the active site, whereas the remainder of the protein as well as the surrounding solvent is treated by a MM method (Fig. 1B). This is intended to combine the accuracy of the QM method with the speed of the MM method. This is accomplished by adding QM and MM energies and forces, avoiding double-counting any interactions.

There are many ways to obtain a valid QM/MM energy function (Senn & Thiel, 2009). The simplest approach is:

$$E_{\mathrm{ME}}^{\mathrm{QM/MM}} = E_1^{\mathrm{QM}} + E_{12}^{\mathrm{MM}} - E_1^{\mathrm{MM}} \tag{1}$$

where E_1^{QM} is the QM energy of the QM system (subsystem 1), E_1^{MM} is the MM energy of the same system, and E_{12}^{MM} is the MM energy of all atoms. Such an energy can be obtained with any combination of QM and MM programs without any change in the code and it is employed in the ONIOM approach (Svensson et al., 1996).

Eq. (1) indicates that all interactions between the QM and MM systems are treated at the MM level. This is appropriate for van der Waals interactions, which are hard to describe accurately with QM, but it is more questionable for electrostatic interactions, which often provide the dominating catalytic effect. Alternatively, the electrostatic interactions between the QM and MM systems can be treated with QM by including the MM point charges in the QM calculation (E_{1,q_2}^{QM}) and turning off the corresponding interactions in the MM calculations by zeroing the charges in the QM system ($E_{12,q_1=0}^{\mathrm{MM}}$ and $E_{1,q=0}^{\mathrm{MM}}$) (Ryde, 1996):

$$E_{\mathrm{EE}}^{\mathrm{QM/MM}} = E_{1,q_2}^{\mathrm{QM}} + E_{12,q_1=0}^{\mathrm{MM}} - E_{1,q=0}^{\mathrm{MM}} \tag{2}$$

This is called electrostatic (electronic) embedding (EE), in contrast to the mechanical embedding (ME) in Eq. (1). It allows the QM system to be polarized by the MM charges, but the MM system is not polarized by the MM charges. The most accurate approach is to allow both the QM and MM systems to be simultaneously polarized (polarized embedding) (Poulsen, Kongsted, Osted, Ogilby, & Mikkelsen, 2001; Söderhjelm, Husberg, Strambi, Olivucci, & Ryde, 2009), but this requires a polarizable MM potential and a QM program that allows for the inclusion of MM polarizabilities.

If there are covalent bonds between the QM and MM systems, so-called junctions, special action is needed. First, the QM system needs to be truncated in a proper way. This can be done in two different ways. One is to truncate the QM system with hydrogen atoms (as in the QM-cluster approach; the H-link-atom approach) or with parametrized boundary atoms (Senn & Thiel, 2009; Von Lilienfeld, Tavernelli, & Rothlisberger, 2005; Zhang, 2006). The alternative is to use frozen localized orbitals at the boundary (Warshel & Levitt, 1976). The localized self-consistent field, suggested by Rivail and coworkers (Ferré, Assfeld, & Rivail, 2002; Théry, Rinaldi, Rivail, Maigret, & Ferenczy, 1994) and later extended and parametrized by Friesner and coworkers (Philipp & Friesner, 1999), places the frozen orbital on the last QM atom, whereas the generalized hybrid orbital method by Gao and coworkers instead introduces a boundary atom with a localized orbital (Gao, Amara, Alhambra, & Field, 1998; Pu, Gao, & Truhlar, 2005). In theory, the localized orbitals should give the more accurate results when properly parametrized, but in practice the two approaches typically give similar results (Nicoll, Hindle, MacKenzie, Hillier, & Burton, 2001). The localized orbital approaches require specialized QM software, whereas the H-link-atom approach can use any QM software. Therefore, most QM/MM studies use the latter approach (Ryde, 2003).

It should be recognized that junctions provide a severe problem for the QM/MM calculations. The H-link-atom approach introduces an incorrect atom at an erroneous position. These problems are further reinforced in EE approaches by the presence of nearby point charges, which give rise to strong artificial interactions (Hu, Söderhjelm, et al., 2011). Therefore, many schemes have been suggested to remove or redistribute charges around the junctions (König, Hoffman, Trauenheim, & Cui, 2005; Lin & Truhlar,

2005; Sherwood et al., 1997). However, we did not find any consistent improvement by any of these approaches in a large-scale test, except that the charge on the atom that is converted to the link atom should be excluded (Hu, Söderhjelm, et al., 2011). It has also been noted that all problems with the link atoms can be corrected at the MM level (Rod & Ryde, 2005b; Vreven et al., 2006). However, in practice, such corrections do more harm than good if the junctions are close to the reactive center (Hu, Söderhjelm, et al., 2011).

ME approaches are not affected by these problems because the electrostatic interactions are evaluated at the MM level in the full system without any junctions. However, they typically employ some sort of QM charges, which have been obtained with link atoms and therefore have to be adapted to the real system in some more or less arbitrary way (Hu, Söderhjelm, et al., 2011). Alternative ME approaches have been developed, eg, with QM charges from a polarized wavefunction, with a proper treatment of polarization, and avoiding instabilities in the determination of QM charges, which give significantly better results than standard ME and EE approaches, especially when the junctions are close to the reactive system (Hu, Söderhjelm, et al., 2011).

Finally, it has often been noted that EE approaches give rise to an overpolarization of the QM system. This has been attributed to the charges close to the link atoms (Senn & Thiel, 2009), but it is also caused by the inconsistent treatment of polarization (the QM system is polarizable, but not the MM system; Hu, Söderhjelm, et al., 2011). Several groups have suggested that the van der Waals parameters around the junctions should be modified (Freindorf, Shao, Furlani, & Kong, 2005; Murphy, Philipp, & Friesner, 2000; Riccardi, Li, & Cui, 2004). However, it seems dangerous to correct a problem in the electrostatics by modifying van der Waals parameters. In practice, the problems are best solved by minimizing the number of junctions and moving them as far away as possible from the reactive site (Hu, Söderhjelm, et al., 2011).

One of the greatest advantages with the QM/MM calculations is that they allow for a thorough interpretation and understanding of the results. Once it has been shown that the QM/MM calculations give reasonable results that reproduce key experimental findings, an analysis of the QM/MM results and components can show how the surroundings affect the structures and reaction energies. Moreover, such an analysis also gives an indication whether the results are reliable and therefore should always be performed. This is thoroughly illustrated by the application in Section 3.1.

The QM/MM approach was introduced by Warshel and Levitt (1976) for energies and by Singh and Kollman (1986) for geometry optimization. QM/MM programs are now available in many QM or MM software, eg, Gaussian (ONIOM; Svensson et al., 1996), Q-site (Murphy et al., 2000), AMBER (Götz, Clark, & Walker, 2014), and CHARMM (Riahi & Rowley, 2014), and also in some independent QM/MM interfaces that combine various QM and MM software, eg, ChemShell (Sherwood et al., 2003) and ComQum (Ryde, 1996; Ryde & Olsson, 2001). In QM/MM applications, it is important to have a versatile MM software with the opportunity of doing MD simulations for structure and solvent equilibration and also with utilities to parameterize nonstandard molecules in the protein—typically the MM calculations are much harder to set up than the QM calculations because they rely on parameterized empirical potentials.

2.4 The Big-QM Approach

One of the prime problems of both the QM-cluster and QM/MM approaches is that the selection of the QM system may bias the results, ie, that if the QM system is extended with more groups, the energies may change significantly. Himo and coworkers have shown that QM-cluster results may depend on the size of the QM system and also the value of the dielectric constant used for the continuum solvation model. However, when 150–200 atoms have been included, the results typically become essentially independent of the dielectric constant, which has been used as a convergence criterion (Hopmann & Himo, 2008; Sevastik & Himo, 2007). On the other hand, we have shown that even with QM systems of 400 atoms, the energies may be unstable: We obtained a difference of over 50 kJ/mol if the QM residues were selected according to their distance to the active site or by QM/MM energy components (Hu, Eliasson, Heimdal, & Ryde, 2009; Sumner et al., 2013). Similar differences have also been found in other studies, eg, a 45 kJ/mol difference in QM-cluster calculations with 300 and 1800 atoms (Sumowski & Ochsenfeld, 2009) and 55 or 27 kJ/mol differences between QM systems of 27 or 135 and 220 atoms, respectively (Liao, Yu, & Himo, 2011).

Several studies have indicated that QM/MM calculations converge faster than QM-cluster calculations with respect to the size of the QM system (Flaig, Beer, & Ochsenfeld, 2012; Sumowski & Ochsenfeld, 2009). On the other hand, other investigations have shown that also QM/MM energies strongly depend on the size of the QM system (Tian, Strid, &

Eriksson, 2011). In fact, we showed that the convergence of QM/MM calculations for a model of [NiFe] hydrogenase was not much better than QM-cluster calculations unless specialized ME approaches were used (Hu, Söderhjelm, et al., 2011).

Our convergence studies of QM-cluster and QM/MM calculations gave clear indications of what groups need to be included in the QM system to give converged results (Hu et al., 2009; Hu, Söderhjelm, et al., 2011): Neutral groups contribute significantly to reaction energies only when within 4.5 Å of the reactive system. Charged groups, on the other hand, influence the energies even at distances of 16 Å. However, it is known from experiments that solvent-exposed charges do not affect pK_a values and other energies (André, Kesvatera, Jönsson, Åkerfeldt, & Linse, 2004). Moreover, we have already emphasized that junctions should be moved away from the reactive site. Based on these results, we suggested the big-QM approach to obtain stable energies for protein reactions (Hu et al., 2013). In this approach, all chemical groups within 4.5–6 Å of the minimal QM system, all buried charged groups in the protein, and two capped amino acids around each residue in the minimal QM system are included in the QM calculations (Fig. 1C). This typically amounts to 600–1000 atoms, for which single-point energies easily can be calculated at the DFT level with SVP or even VTZ basis sets. Liao and Thiel (2013) have also argued for large QM systems (408–657 atoms), selected by a charge deletion approach (ie, single-point QM/MM calculations with the charges of one MM residue deleted).

The big-QM energies can be calculated in vacuum, with a point-charge model, or in a continuum solvent. The former can then be corrected by an ME-QM/MM term, according to Eq. (1), whereas the EE approach (Eq. 2) can be used for the energy obtained with the point-charge model. In practice, all three approaches give similar results, within 14 kJ/mol (Hu et al., 2013). A point-charge model typically gives the fastest convergence in the QM calculations and is therefore recommended, especially as all junctions are far from the reactive system.

The effect of the size of the QM system in the geometry optimizations has also been examined (Sumner et al., 2013). It was shown that QM/MM optimizations give much more stable and reliable structures than QM-cluster optimizations with various sets of fixed atoms. The raw QM/MM energies showed a quite large variation with the size of the QM system (up to 70 kJ/mol), but converged after the inclusion of 6–13 residues. The big-QM energies, on the other hand, were very stable, with variations of less than 15 kJ/mol, mainly depending on residues involving

junctions. Therefore, QM/MM optimizations can employ quite small QM systems if energies are evaluated by the big-QM approach.

2.5 QM/MM Free-Energy Calculations

The third serious issue with QM/MM calculations comes from the local-minima problem. A minimization typically converges to the closest local minimum, which is not necessarily the global minimum of the system. There are no optimization methods that always find the global minimum (except an exhaustive systematic search). In practice, it is normally not necessary that all groups reside in their global minimum, but it is essential that all distant groups remain in the same local minimum throughout a reaction sequence; otherwise the energy will be blurred by irrelevant energy components. For example, the total energy may change by ~20 kJ/mol if a single water molecule at the periphery changes its hydrogen-bond pattern. For small QM-cluster models, you can see by the eye whether all states in a reaction mechanism belong to the same local minimum. However, for a solvated protein, this is impossible—there are thousands of atoms and essentially an infinite number of local minima.

This is a very serious problem for QM/MM calculations and several approaches have been suggested to detect and avoid it (Senn & Thiel, 2009). One way is to run calculations forth and back between each pair of states in a reaction mechanism until the energy differences are constant. This is tedious and time consuming, especially if the mechanism contains many states. Another way is to keep the relaxed MM system as small as possible. However, this introduces the risk that important changes in the surroundings are missed, in particular the dielectric relaxation of the surroundings in effect of changes in the charge distribution of the active site. Moreover, even a rather small part of the protein may have many local minima.

A more common approach is to perform a MD simulation and then calculate QM/MM structures for several (3–10) snapshots from the simulations (Hu et al., 2013; Lonsdale, Harvey, & Mulholland, 2010; Metz & Thiel, 2009). This shows possible variations of the energies and indicates how certain the results are. However, it is not obvious how the results should be averaged to give reliable activation barriers (Cooper & Kästner, 2014).

The local-minima problem is caused by the fact that you consider minimized structures. The proper way to cure it is to calculate free energies, which involves sampling and averaging over all relevant thermally accessible

structures. Several strict methods are available to obtain valid free energies, in particular free-energy simulations (FES), based on MD or Monte Carlo simulations and calculating the free energies by exponential averaging, thermodynamic integration, or Bennett acceptance ratio (Bennett, 1976; Kirkwood, 1935; Zwanzig, 1954). Such calculations are today routinely performed at the MM level, often giving quite accurate results (Hansen & van Gunsteren, 2014; Mikulskis, Genheden, & Ryde, 2014).

FES calculations can be performed also with QM/MM if an SEQM method is employed (Reddy & Erion, 2007; Senn & Thiel, 2009; Senn, Thiel, & Thiel, 2005). However, the rather poor accuracy of the SEQM methods makes such an approach risky, especially for transition states with their complicated electronic structure (Heimdal & Ryde, 2012).

QM/MM FES with DFT or high-level QM methods are rare and restricted to short simulation times owing to the large computational effort (Senn & Thiel, 2009; VandeVondele & Rothlisberger, 2002). Therefore, methods have been developed to avoid the costly QM/MM simulations. They are typically based on the thermodynamic cycle in Fig. 2. The aim is to estimate the free-energy difference between two states, A and B, at the QM/MM level, but this is computationally prohibitive. However, at the MM level this free energy can easily be obtained by FES methods. With these MM simulations, we can also estimate the free-energy difference of going from MM to QM/MM for the A and B states, completing the thermodynamic cycle. Such reference-potential methods were first suggested by Luzhkov and Warshel (1992) and they have been reinvented several times (Iftimie, Salahub, Wei, & Schofield, 2000; Wood, Yezdimer, Sakane, Barriocanal, & Doren, 1999), eg, the QM/MM thermodynamic cycle perturbation approach (QTCP; Rod & Ryde, 2005a, 2005b).

The problem with such an approach is that the MM → QM/MM perturbations need to converge in a single step if QM/MM simulations should

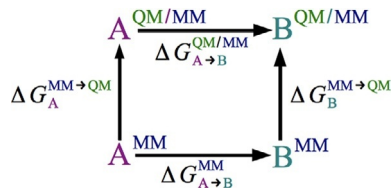

Fig. 2 The thermodynamic cycle that is the basis of the reference-potential methods and QTCP, showing that $\Delta G^{QM/MM}_{A\to B} = \Delta G^{MM}_{A\to B} - \Delta G^{MM\to QM}_{A} + \Delta G^{MM\to QM}_{B}$ (Luzhkov & Warshel, 1992; Rod & Ryde, 2005a).

be avoided, although FES studies (eg, the A → B perturbation at the MM level) typically need to be divided into 10–40 small steps to ensure proper convergence. This is normally solved by keeping the QM system fixed, which ignores differences in the QM and MM energy function for the internal degrees of freedom of the QM system (Rod & Ryde, 2005b). This means that the entropy of the QM system is omitted, but it can be estimated from QM frequencies of the isolated QM system. This is often a valid approximation for enzyme reactions, but it is more questionable when studying the binding of a ligand to a protein. Therefore, there has recently been quite some interest to obtain full QM/MM free energies with a flexible QM system, eg, by employing MD simulations at the SEQM/MM level or with a tailored MM potential (Heimdal & Ryde, 2012), or by employing the non-Boltzmann Bennett acceptance ratio (NBB) approach (König & Boresch, 2011). Recently, we have shown that converged QM/MM binding free energies require ~700,000 QM calculations if interaction energies are used and that exponential averaging with the cumulant expansion gives both a faster convergence and requires fewer QM calculations than NBB (Olsson, Söderhjelm, & Ryde, 2016).

Alternative approaches have been suggested, involving some restricted QM/MM MD simulations. In the paradynamics approach, one short QM/MM MD simulation is performed for each state and the free energy is estimated by the linear response approximation (Plotnikov, Kamerlin, & Warshel, 2011). Woods, Shaw, and Mulholland (2015) instead used the Metropolis–Hastings Monte Carlo approach to reduce the number of QM/MM calculations needed.

Several more approximate methods have also been developed to obtain QM/MM free energies. In the QM/MM-FE (free energy) approach (Hu & Yang, 2008; Zhang, Liu, & Yang, 2000) the MM → QM/MM perturbations are replaced by single-point MM → QM/MM extrapolations (ie, $\Delta E_{QM/MM} - \Delta E_{MM}$ for the QM/MM structures of the two states A and B). Even if this is a quite severe approximation, it has been shown to reproduce QTCP free energies within 5–9 kJ/mol for two sets of proton-transfer reactions (Kaukonen et al., 2008a). However, this approach is still quite time consuming as it involves FES at the MM level with extensive MD sampling.

The MM/PBSA (molecular mechanics combined with Poisson–Boltzmann and surface area solvation) is a popular approach to estimate binding free energies (Genheden & Ryde, 2015; Kollman et al., 2000). In this approach, free energies are estimated from internal, electrostatic,

and van der Waals energies, calculated at the MM level, combined with PB or GB electrostatic continuum solvation energies, nonpolar solvation energies from the solvent-accessible surface area, as well as entropies from vibrational frequencies obtained at the MM level. Several groups have developed QM/MM-PBSA approaches, in which the MM energies are replaced by QM/MM calculations, keeping the solvation-free energies and entropies (Gräter, Schwarzl, Dejaegere, Fischer, & Smith, 2005; Kaukonen, Söderhjelm, Heimdal, & Ryde, 2008b; Retegan, Milet, & Jamet, 2009; Ryde & Söderhjelm, 2016). We have shown that such an approach reproduces QTCP proton-transfer reaction free energies within 7–13 kJ/mol even if only minimized QM/MM structures are used, which is appreciably better than the raw QM/MM energies (Kaukonen et al., 2008b). Such energies can be obtained from QM/MM structures with a minimal effort.

2.6 Combination with Experiments

QM/MM investigations are typically based on protein crystal structures. Unfortunately, crystal structures cannot be used directly to extract QM energies, owing to systematic errors in both the QM and experimental methods (Ryde et al., 2002). Therefore, most QM/MM projects involve reoptimization of crystal structures. Outer atoms are then typically fixed at the original crystal structure, whereas the active site and surrounding residues are freely optimized. Thereby, the latter atoms may move significantly away from the true structure, whereas the positions of the outer atoms may be biased by uncertainties in the crystal structure. A more satisfactory approach would be to restrain all atoms toward the experimental raw data, ie, the crystallographic structure factors.

We developed such a quantum-refinement approach in 2002 (Ryde et al., 2002). Standard crystallographic refinement is already in the form of an optimization in which the deviation between the experimental structure factors and those calculated from the current coordinates should be minimized (measured by the crystallographic R factor or more sophisticated statistical measures, E_{12}^{cryst}; Kleywegt & Jones, 1997). Moreover, the crystallographic data are typically supplemented by a MM-like energy function to ensure that bond lengths and angles are chemically reasonable:

$$E^{\text{X-refine}} = w_{\text{A}} E_{12}^{\text{cryst}} + E_{12}^{\text{MM}} \tag{3}$$

The w_{A} weight factor is needed to because E_{12}^{cryst} typically is unit-less, whereas the MM term is in energy units. w_{A} gives the relative weight of

the crystallographic and MM data and it is normally determined by requiring that the crystallographic and MM forces during a short MD simulation are of a similar magnitude (showing that standard crystallographic structures actually are 50% theoretical).

QM can be introduced into this function, by replacing the MM potential for a small, but interesting, part of the protein by QM calculations, giving the energy function (Ryde et al., 2002):

$$E^{\text{QM-refine}} = w_A E_{12}^{\text{cryst}} + E_1^{\text{QM}} + E_{12}^{\text{MM}} - E_1^{\text{MM}} \quad (4)$$

We have shown that such an approach works well and that it can improve crystal structures locally, because DFT calculations give structures with a better accuracy than low- and medium-resolution crystal structures (Ryde, 2007; Ryde & Nilsson, 2003). Moreover, it can be used to deduce the protonation state of metal-bound ligands in crystal structures, although the protons are not explicitly seen (Nilsson & Ryde, 2004), or to discriminate between alternate structural interpretations of intermediates in the reaction mechanism (Heimdal, Rydberg, & Ryde, 2008; Hersleth, Ryde, Rydberg, Görbitz, & Andersson, 2006; Ryde, Hsiao, Rulíšek, & Solomon, 2007; Söderhjelm & Ryde, 2006). However, the application of the method to other metalloproteins has been hampered by the fact that metal sites are often partly reduced during the data collection, so that the raw data describes a mixture of structures (Rulíšek & Ryde, 2006; Söderhjelm & Ryde, 2006). Quantum refinement can detect such photoreduction, but no accurate models can be obtained. The method has recently been extended also to neutron crystallographic data (Manzoni, Oksanen, Logan, & Ryde, 2016).

A similar approach can be used also for NMR data: The raw data are different, but standard refinement still involves optimization of an energy function containing one term for the agreement between experimental data and the coordinates and a standard MM energy function (Cavanagh, Fairbrother, Palmer, & Skelton, 1996). Therefore, an energy function of the same type as the one in Eq. (4) can be used to introduce QM/MM calculations for an interesting part of the protein (Hsiao, Drakenberg, & Ryde, 2005). This was tested for two calcium-binding sites in protein S, with strongly improved structures compared to standard NMR refinement.

Finally, we have also developed methods to combine QM-cluster or QM/MM calculations with EXAFS (extended X-ray absorption fine structure) structures (Hsiao, Tao, Shokes, Scott, & Ryde, 2006). This was more

complicated because the standard EXAFS approach gives only information of a few metal–atom distances. This approach has been used to study sitting-atom intermediates in the metallation of porphyrins, to decide the structure of the peroxide adduct in multicopper oxidases and of the oxygen-evolving complex in photosystem I (Hsiao & Ryde, 2006; Li, Siegbahn, & Ryde, 2015; Li, Sproviero, Ryde, Batista, & Chen, 2013; Ryde et al., 2007).

3. APPLICATIONS

In this section, I give a few examples from our QM/MM studies, which have led to our current QM/MM approach or illustrate some of the important aspects of QM/MM investigations.

3.1 Energy Components in Glutamate Mutase

In the first example, I want to illustrate how QM/MM calculations can be interpreted using energy components to give a complete understanding of an enzyme reaction and the catalytic effect (Jensen & Ryde, 2005). Glutamate mutase is a bacterial enzyme that catalyzes the conversion of glutamate to L-*threo*-methylaspartate (Marsh, 2000). It is a radical reaction, initiated by the coenzyme adenosyl cobalamine (AdoCbl), a vitamin B_{12} derivative. AdoCbl contains a Co ion in the center of a corrin ring. In the resting state, Co is in the +III state, forming a Co–C bond to the C5′ methylene group of ribose moiety of adenosine. This bond is relatively weak and can be homolytically cleaved to Co(II) and an adenosyl radical.

In gas phase, the calculated homolytic bond dissociation energies (BDE) of AdoCbl and the related coenzymes RibCbl and MeCbl (with ribose or a methyl group bound to Co) at the BP86/SVP level of theory are 143–156 kJ/mol (Jensen & Ryde, 2005), in reasonable agreement with experimental estimates of 126–155 kJ/mol (Hay & Finke, 1986; Martin & Finke, 1992). However, QM/MM calculations indicated that in the protein, the BDE is only 13 kJ/mol (Jensen & Ryde, 2005), in accordance with experimental estimates of an equilibrium constant close to unity (Padmakumar, Padmakumar, & Banerjee, 1997). Our aim is to understand this difference in BDE.

The BDE is calculated with respect to the infinitely separated products. However, in the protein, the two moieties cannot dissociate to a Co–C distance of more than ~3.5 Å. At this distance, there are still significant interactions between the two moieties and the BDE is only 122 kJ/mol. This has been termed the cage effect (Dölker, Maseras, & Siegbahn, 2004) and it

explains 20 kJ/mol of the catalytic effect. The QM/MM energy difference between the Co(III) and Co(II) states in the protein is 8 kJ/mol.

The latter energy can be divided into the QM and MM energy components, according to Eq. (2), which amount to 33 and 25 kJ/mol, respectively. E_{1,q_2}^{QM} is the QM energy, including a point-charge model of the surroundings. The corresponding energy without the point-charge model is 67 kJ/mol and the difference (34 kJ/mol) represents the direct electrostatic effect of the surroundings. All these energies have been calculated for QM/MM structures. Using instead geometries optimized in vacuum, we get the gas-phase BDE, but still with a Co–C distances of 3.5 Å for the dissociated state, 122 kJ/mol. The difference (56 kJ/mol) represents the geometric effect of the protein. Consequently, we have shown that the catalytic effect of the protein (143 − 8 = 135 kJ/mol) can be divided into MM, electrostatic, geometric, and cage effects of 25, 34, 56, and 20 kJ/mol, respectively (Fig. 3).

Each of these effects can be further understood: All MM terms are pairwise decomposable, so they can be completely divided into contributions from the bonded, van der Waals, and electrostatic interactions between

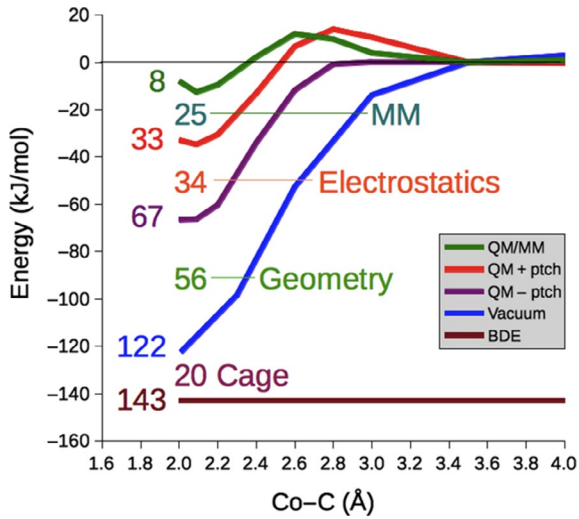

Fig. 3 Energy components for glutamate mutase (Jensen & Ryde, 2005). The *colored numbers* to the *left* are the energies for the five curves at Co–C = 2.0 Å (using the energy at Co–C = 3.5 Å as the reference). The *numbers* to the *right* are the pairwise differences between these numbers. All energies are in kJ/mol. The components are further explained in the main text.

two atoms (three or four atoms for angles and dihedrals). However, the number of terms rapidly becomes overwhelmingly large, so it is normally best to group them. For example, the MM term can be divided into contributions from the QM system (subsystem 1), the optimized part of the MM system (subsystem 2), and the fixed part of the MM system (subsystem 3). The pairwise contributions from these three subsystems are shown in Fig. 4. All terms are dominated by the electrostatic and van der Waals contributions. It can be seen that three of these contributions are negligible: 1–3 (owing to the large distance between these two subsystems), 2–3 (by chance), and 3–3 (because subsystem 3 is fixed). The other three terms are rather small and of a similar magnitude, 5–11 kJ/mol. This is good because it is in these terms (especially 2–2) the local-minima problem is expected. Therefore, the MM term should always be investigated and if it is large, its components should be thoroughly checked.

The direct electrostatic term is normally the dominating effect of the protein. It comes from QM calculations, making it somewhat harder to interpret. It can be residue-wise decomposed by repeating the QM calculation with the point charges of one residue deleted (Liao & Thiel, 2013). However, this is somewhat tedious and time consuming, and it still ignores the polarization effects (ie, all residue components will not add up to the total effect). Normally, the electrostatic effects are similar if instead ME-QM/MM is used, ie, if the electrostatic effects are calculated with MM (Eq. 1) with a QM charge model of the QM system (Ryde, 1996). Then, the electrostatic components can be obtained in a single (cheap)

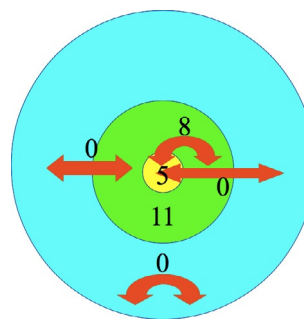

Fig. 4 MM energy components in kJ/mol (Jensen & Ryde, 2005). The *yellow* (*white* in the print version), *green* (*light gray* in the print version), and *cyan* (*light gray* in the print version) *circles* represent the QM, relaxed MM, and fixed MM systems, respectively, and the energy components are the energies within or between the various systems, as illustrated by the *arrows*.

calculation. Fig. 5 shows the major residue contributions from such an analysis of glutamate mutase, indicating that it is mainly charged residues around the adenosyl moiety that affect the reaction energies.

The geometric term is unusually large in this enzyme. It can be interpreted by calculating strain energies in different parts of the QM system, ie, the energy difference between the molecule optimized in vacuum and in the protein. As can be seen in Fig. 6, the strain energy is dominated by

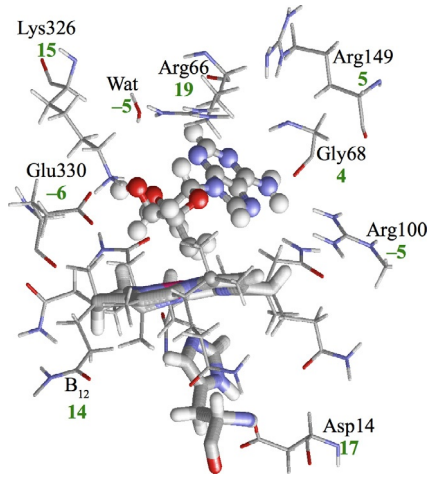

Fig. 5 Residues with large electrostatic energy components (*numbers* in *green* (*gray* in the print version); kJ/mol) for the catalytic effect of glutamate mutase (Jensen & Ryde, 2005). The axial His ligand and the corrin ring are shown in *thick sticks*, whereas the adenosine moiety is shown with *balls and sticks*.

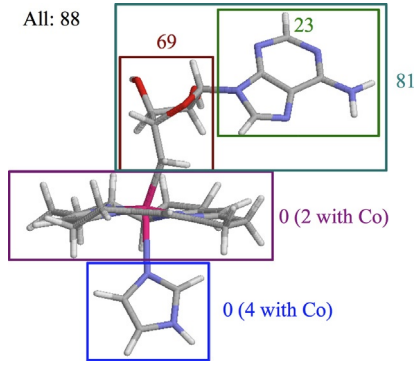

Fig. 6 Strain energy in various moieties of the AdoCbl coenzyme in kJ/mol (Jensen & Ryde, 2005).

contributions from the adenosyl moiety and especially the ribose ring. Further calculations with systematically removed constraints to the QM/MM structure show that it is especially the Co–C–C angle that is constrained, by ~40 kJ/mol.

Next, we can turn the attention to the surrounding protein and ask how it constrains the geometry. By repeating the QM/MM optimization with zeroed charges of the protein, we showed that the electrostatic effect on the geometry is rather small, ~24 kJ/mol (Jensen & Ryde, 2005). On the other hand, if also the van der Waals contributions were turned off, the catalytic effect of the protein almost disappeared. Finally, the residues that are most important for the geometric effect could be determined by repeating the QM/MM optimizations with some residues around the active site mutated to Gly. The results in Fig. 7 show that the effect comes mainly from two residues (Lys326 and Glu330) and one of the side chains of the cobalamine coenzyme, although the effects are strongly cooperative. All important groups form hydrogen bonds to the ribose moiety of the coenzyme, showing that this part is needed as a handle. This was also confirmed by repeating the QM/MM calculations with the MeCbl and RibCbl coenzymes—the latter still gave a catalytic effect of 109 kJ/mol, whereas it was reduced to only 42 kJ/mol for MeCbl.

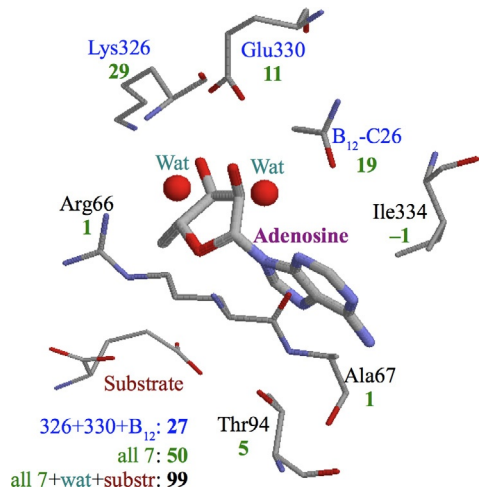

Fig. 7 Decrease in catalytic effect when various residues are mutated to Gly in kJ/mol (Jensen & Ryde, 2005). The cooperative effect of simultaneously mutating three or the seven shown residues is given in the *lower left corner*, together with the effect of mutating the same seven residues and deleting two water molecules and the substrate (also shown in the figure).

Finally, it should be noted that several aspects of this reaction are controversial, both regarding what DFT functional should be used (Jensen & Ryde, 2003; Kozlowski et al., 2012; Siegbahn, Blomberg, & Chen, 2010) and the importance of geometric vs electrostatic effects (Jensen & Ryde, 2009; Kwiecien, Khavrutskii, Musaev, Morokuma, Banerjee, & Paneth, 2006; Román-Meléndez, von Glehn, Harvey, Mulholland, & Marsh, 2014; Sharma, Chu, Olsson, & Warshel, 2007). The prime problem is that there are no reliable crystal structures with an intact Co(III)–C bond, owing to photoreduction of the structures. Still, this example nicely shows that QM/MM results are not restricted to geometries and energies; much more information can be obtained from the QM/MM energy components and additional test calculations.

3.2 Mo Oxo-Transfer Enzymes

The second application emphasizes the accuracy of the QM calculations and how it depends on the size of the QM system. A large number of oxo-transfer enzymes employ molybdenum in the active site. For example, DMSO reductase converts dimethyl sulfoxide (DMSO) to dimethyl sulfide and the active-site Mo ion is concomitantly oxidized from Mo(IV) to Mo(VI)═O (Hille, Hall, & Basu, 2014). At least six research groups have performed QM-cluster calculations on this enzyme and all have agreed on a two-step reaction mechanism with an activation enthalpy of 68–80 kJ/mol (Ryde, Dong, Li, Andrejic, & Mata, 2016), ie, close to the experimental estimate of 63 kJ/mol (Cobb, Conrads, & Hille, 2005).

However, we found that the reaction is unusually sensitive to the QM method (Li, Andrejic, Mata, & Ryde, 2015; Li, Mata, et al., 2013). The activation enthalpy increased by 74 kJ/mol if calculated with TZP instead of SVP basis sets, whereas a further increase to QZP changed it by 9 kJ/mol. The effect of dispersion for the QM system (estimated with DFT-D2) was large, eg, 33 kJ/mol for the activation energy. Moreover, this energy needs to be balanced by the nonpolar solvation energies (which also contain a dispersion component), especially when a molecule binds or dissociates from the active site (Ryde, Mata, & Grimme, 2011). Unfortunately, such terms are problematic for enzyme active sites, for which it has to be decided whether there is a cavity before the ligand binds and if it is filled with solvent or not. These terms also contribute by ~38 kJ/mol to the activation enthalpy.

Furthermore, different DFT functionals gave widely different activation and reaction energies, owing to the change in the oxidation states of Mo and the substrate: The activation enthalpy was low (23–34 kJ/mol) with pure functionals, but much higher with hybrid functionals, 46–139 kJ/mol, increasing with the amount of HF exchange. To know what results to trust, we performed LCCSD(T0) calculations, extrapolated to the complete-basis-set limit (CBS). These showed that no single functional gave reliable results—instead B3LYP gave the best activation enthalpy, whereas the pure functionals gave better results for the other states. The DFT results were also much closer to the LCCSD(T0) reference if a DFT-D correction was included, showing that dispersion corrections clearly improve DFT results. Our final results, based on the LCCSD(T0)/CBS energies and including relativistic, zero-point-energy, thermal, and solvation corrections (both polar and nonpolar terms), reproduced experimental results both for the enzyme reaction and for an inorganic model (Li, Mata, et al., 2013; Li, Sproviero, et al., 2013).

Interestingly, our calculations gave nearly the same activation enthalpy as the previous computational studies, although our estimate included dispersion and nonpolar solvation corrections of 71 kJ/mol that were not considered before. The reason is that the previous studies either used a method (BP86) or a basis set (SVP) that gave a too low barrier. Thus, all previous studies reproduced the experimental results for the wrong reason. On the other hand, they gave widely different reaction energies (–50 to –131 kJ/mol) because no experimental results were available for this energy.

A similar strong method dependence was obtained also for the sulfite oxidase reaction ($SO_3^{2-} + Mo^{6+}{=}O \rightarrow SO_4^{2-} + Mo^{4+}$), which we solved with LCCSD(T0)/CBS calculations and the same type of energy corrections (Van Severen et al., 2014). However, in this case QM-cluster calculations gave a too high activation energy even in a water-like continuum solvent, owing to extensive Coulombic repulsion between the negatively charged active site and substrate. In the protein, this is compensated by three Arg residues. This is best studied by QM/MM methods enzyme (Caldararu, Andrejic, Cioloboc, Mata, & Ryde, 2016). It turned out that a large QM system was needed to obtain stable structures and energies, including the three Arg residues, as well as five additional groups and two water molecules (164 atoms; Fig. 8A). With such a model, a reasonable activation barrier could be obtained (~60 kJ/mol). Moreover, this large and carefully selected QM system gave almost identical energies (within 4 kJ/mol) to those of an 805-atom big-QM system (Fig. 8B).

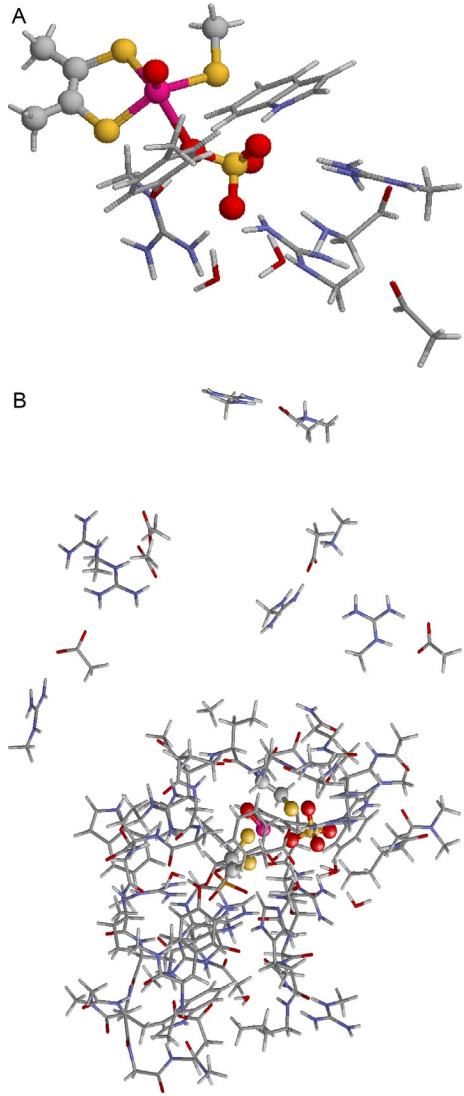

Fig. 8 The QM system needed for a proper QM/MM study of sulfite oxidase (A) and the corresponding big-QM system (B; Caldararu et al., 2016).

These studies show several important aspects of QM studies of proteins:

- Always test both pure and hybrid DFT functionals for your reaction of interest and use LCCSD(T0)/CBS or similar methods as a reference if the results differ.
- Basis sets of QZP quality should be used to ensure convergence of the energies.

- DFT energies should always be supplemented by dispersion corrections.
- Corrections for relativistic, zero-point-energy, thermal, and solvation effects should always be included.
- Solvation energies should include also the nonpolar terms, especially when molecules bind or dissociate. These terms are quite problematic in QM-cluster calculations of an enzyme active site.
- The QM system in QM/MM studies should be selected to give stable structures and energies, including all significant movements of the surroundings. Such a properly chosen QM system often gives energies close to those obtained with the big-QM approach.

3.3 Blue Copper Proteins and Multicopper Oxidases

Blue copper proteins are small electron carriers. They contain a Cu site, typically with one Cys, two His, and possibly one or two additional weaker ligands, Met or a carbonyl group (cf. Fig. 9A). The structure is normally trigonal with the three strong ligands in an approximate plane and 0–2 axial ligands. The Cu ion alternates between the +I and +II oxidation states, but both states give similar structures, mainly owing to the properties of the Cu–Cys bond (Ryde, Olsson, & Pierloot, 2001). On the other hand, the redox potential can vary quite extensively, from 0.18 to over 1 V (Liu et al., 2014). Many computations studies of the redox potentials of blue copper proteins have been published (Hadt et al., 2012; Olssson, Hong, & Warshel, 2003; Van den Bosch et al., 2005).

It is extremely demanding to obtain stable QM/MM redox potentials for metal sites in proteins, as is illustrated by the results in Table 1 (M.-C. Van Severen, M. Kaukonen, & U. Ryde, unpublished results): QM/MM potentials depend strongly on the oxidation state for which the MM system is optimized with differences of 1.5 V, even when averaged over five independent structures. Fortunately, this variation is reduced to only 0.1 V with QTCP. On the other hand, the results then depend on the treatment of long-range solvation effects, outside the simulated system (40 Å radius): Born/Onsager, Ewald, PB, and GB give results that vary from –2.1 to +0.7 V. This variation can be reduced to 0.2–0.8 V by neutralizing all charged groups on the protein surface. However, the largest problem is that the results also depend on the size of the QM systems: For example, we got average redox potentials at the QM/MM level of 1.3, –1.0, 1.2, 0.5, and 0.0 V using increasingly larger QM systems of 33, 124, 170, 197, and 577 atoms (shown in Fig. 9). Apparently, it is very hard to get converged QM/MM absolute redox potentials. Another problem is that different DFT methods (in particular pure and

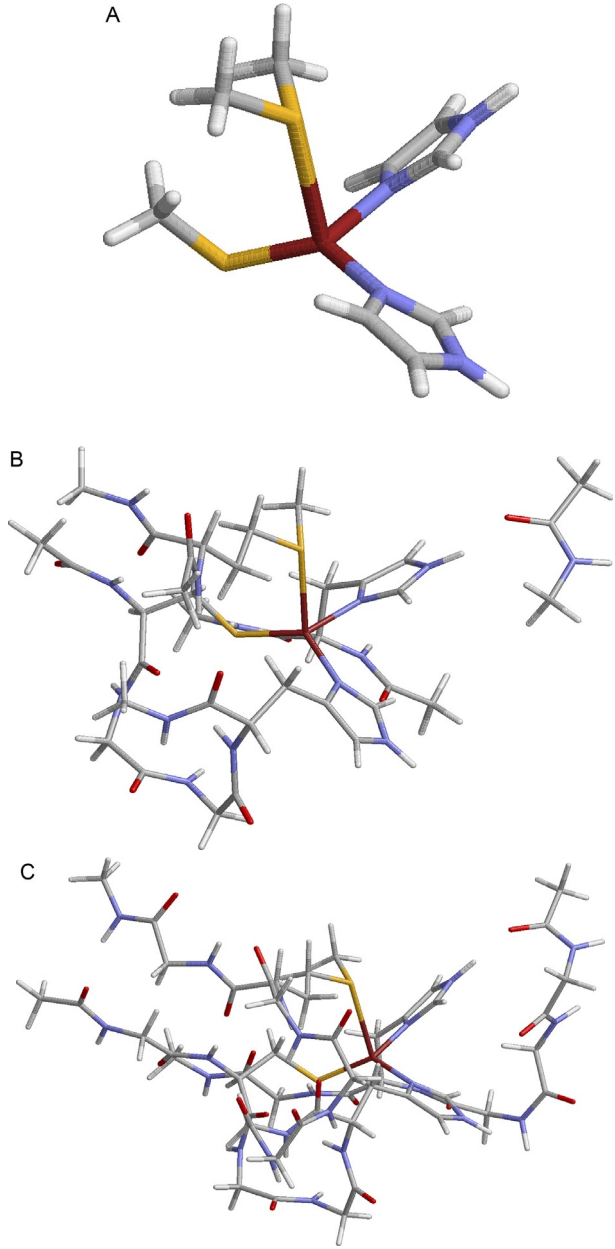

Fig. 9 QM systems of five sizes for plastocyanin: (A) minimal model, (B) junctions moved one residue away and including hydrogen bonds to the first-sphere ligands, (C) junctions moved two residues away, (D) junctions moved three residues away, and (E) a big-QM model.

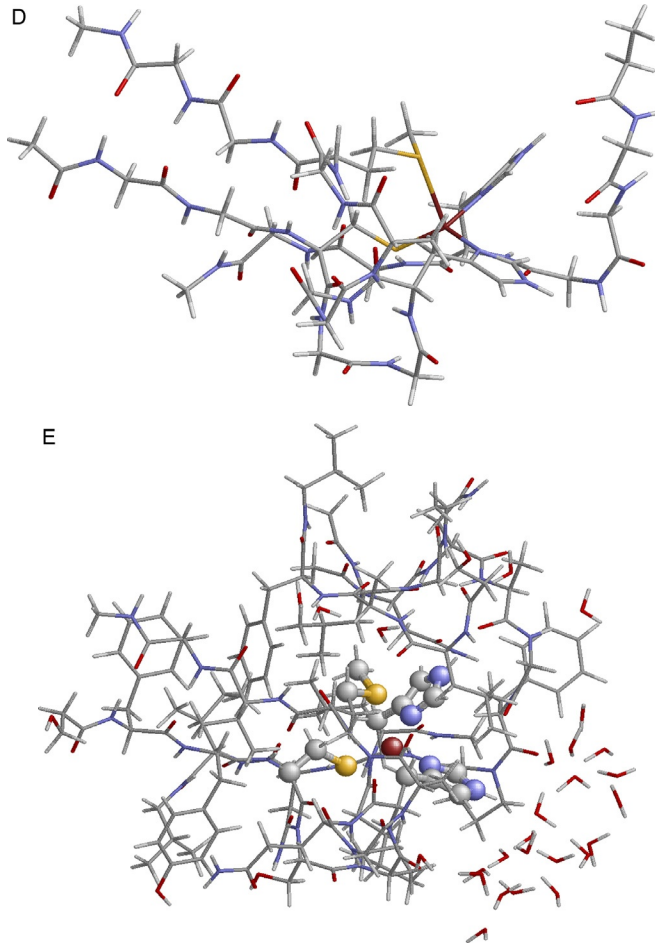

Fig. 9—Cont'd

hybrid functionals) often give redox potentials that differ by 50–100 kJ/mol (Chen, Liu, VandeVondele, Sulpizi, & Sprik, 2014; Heimdal et al., 2011; Li, Mata, et al., 2013).

The former problem can partly be reduced by not calculating absolute potentials, but instead study the internal electron transfer within the protein. Such calculations require the extension of QM/MM to the use of two separate QM systems (Blumberger, 2010; Hu, Farrokhnia, et al., 2011). Such an approach was tested for the multicopper oxidases, which contain both a blue copper site and a trinuclear Cu cluster, in which O_2 is reduced to water. We studied the transfer of electrons between these two sites (~14 Å distance) for

Table 1 The Redox Potential ($E°$ in V) of Plastocyanin, Calculated with Either QM/MM or QTCP (the Latter with Four Different Methods to Treat Long-Range Electrostatics), with Different QM Systems (a–e, Shown in Fig. 9), and with the Protein Surface Charges Neutralized or not

QM System	Method	Neutral Protein	$E°$	$\Delta E°$
a	QM/MM	No	−3.18	1.50
a	QTCP/Born	No	−2.10	0.12
	QTCP/Ewald		0.70	0.07
	QTCP/GB		0.46	0.02
	QTCP/PB		−1.18	0.08
a	QTCP/Born	Yes	0.23	0.32
	QTCP/Ewald		0.76	0.17
	QTCP/GB		0.57	0.31
	QTCP/PB		0.46	0.29
a	QM/MM	Yes	1.28	2.00
b	QM/MM	Yes	−0.95	1.48
c	QM/MM	Yes	1.24	1.00
d	QM/MM	Yes	0.49	0.95
e	QM/MM	Yes	0.02	1.81

$\Delta E°$ is the difference in reduction potentials of structures with the MM system optimized for the reduced or the oxidized state. All results are averages over 10 QM/MM structures (Van Severen et al., unpublished data).

several putative intermediates in the O_2 reduction cycle with both QTCP and QM/MM-PBSA (Li, Farrokhnia, et al., 2015). The results still showed some variation with the continuum solvation model, but the QTCP results obtained with a Born/Onsager term or with Ewald summation gave consistent results. Since the redox potential of the blue copper site in this protein is known from experiments, the absolute potentials of the trinuclear site can be estimated. Moreover, the results showed a significant communication between the two redox sites. For example, the redox potential of the blue copper site varied by at least 0.18 V depending on the state of the trinuclear Cu site.

Similar methods can be used also to estimate acidity constants in proteins. Again, calculations of absolute potentials are extremely hard to converge. However, more stable estimates can be obtained by studying the proton

transfer to a solvent-exposed reference site, especially if solvent-exposed charges are neutralized (Li, Farrokhnia, et al., 2015). The results showed some dependence on the choice of the reference site (±4 pK_a units), but this dependence disappears if the relative acidity is considered. QM/MM-PBSA and -GBSA results differed by over 20 pK_a units, but the QTCP results were quite stable and intermediate between the former two results. By considering the acidity constants of the various putative intermediates, the most stable structure at pH 7 could typically be determined. Combining these results with redox potentials, isomerization energies as well as O_2-binding and water-dissociation energies allowed for the suggestion of a full reaction cycle of the multicopper oxidases. In most cases, the various states were well separated in energies, allowing for a reliable prediction of the most stable states, even if the accuracy of the QTCP predictions was not better than ~30 kJ/mol.

These applications have illustrated the problems encountered when the net charge of the simulated system changes and how they can be solved or reduced by studying only internal electron or proton transfer.

4. SUGGESTED APPROACH FOR QM/MM INVESTIGATIONS

Based on the methods and examples described in the previous sections, as well as our other investigations of protein structure, function, and mechanisms (Delcey et al., 2014; Dong & Ryde, 2016; Hedegård et al., 2015; Heimdal et al., 2011; Hu, Farrokhnia, et al., 2011; Hu et al., 2013; Kaukonen et al., 2008a, 2008b; Li, Farrokhnia, et al., 2015; Rod & Ryde, 2005a, 2005b; Söderhjelm & Ryde, 2006; Sumner et al., 2013), I suggest the following approach for a QM/MM investigation of a protein:

1. Optimize the geometries of the various intermediates and transition states with EE-QM/MM, using a quite small QM system, a pure functional with DFT-D3 dispersion corrections, and an SVP basis set, eg, TPSS-D3/def2-SV(P). This should be done with both the MM system fixed at the crystal structure and the MM system relaxed.
2. Calculate single-point QM energies with the same QM system and a point-charge model using a QZP basis set and both pure and hybrid functionals (eg, B3LYP/def2-QZVP). If the reaction involves anions, diffuse basis sets should be used.
3. Run big-QM calculations with a pure functional and an SVP or (if it can be afforded) a TZP basis set (we typically use TPSS/def2-TZVP), using a point-charge model of the surroundings. To this energy, the

MM correction in Eq. (2) and a DFT-D3 correction should be added, the latter using Becke–Johnson damping (Grimme, Ehrlich, & Goerigk, 2011) and including three-body terms (Sure & Grimme, 2015). The big-QM system should include all residues within 4.5–6 Å of the minimal QM system, all buried charged groups in the protein, as well as two capped residues around each residue of the minimal QM system.

4. Calculate QTCP free energies at the same level and with the same QM system as the geometry optimization. These calculations employ structures with a fixed MM system.
5. Calculate thermal and entropy corrections for the isolated QM system, freely optimized in vacuum (the same QM system and method as in the geometry optimization). In parallel, the reaction should be fully characterized by a QM-cluster study, to get a feeling of the reaction and to estimate the catalytic effect of the protein.
6. If pure and hybrid DFT functionals give significant different results, high-level QM calculations, eg, LCCSD(T0)/CBS, should be used to decide which is most reliable.
7. If relaxation of the MM system significantly changes the structures and energies, the approach needs to be partly modified. The first step is to optimize the MM system for one state and employ this MM structure for all the other states. If the results still depend on whether the MM system is relaxed or not, or if the results depend on the state for which the MM system is relaxed, the QM system needs to be extended until the energies are stable.
8. Currently, our final energies are obtained from the big-QM energies (step 3), extrapolated with the single-point energies in step 2 or corrected with the LCCSD(T0) in step 6, and including zero-point energies, as well as thermal and entropy corrections from step 5, and a free-energy correction from the QTCP calculations in step 5 (Dong & Ryde, 2016). However, it would be more satisfactorily if the QM system in all calculations could be extended to the point that it reproduces the big-QM results. Then, the big-QM calculations would become superfluous (besides guiding the size of the QM system) and the energies could be based on the QTCP free energies (including the DFT-D3 correction in step 2), extrapolated to the large basis set, and including the thermal corrections from step 5.
9. The results should be checked and interpreted by considering various energy components: First, it is divided into MM, electrostatic, and

geometric components. Next, each of these components can be understood by further calculations, as was described in Section 3.1. EE and ME results should be compared, as well as the effect of relaxation of the protein.

10. Preferably, all these calculations should be repeated on a number of snapshots (3–10) from an MD simulation. The QTCP calculations are expected to reduce the starting-point dependence (cf. Table 1), but probably not completely.
11. If QTCP energies are hard to calculate (eg, for the binding of a ligand from the solution) or more approximate results can be accepted, QM/MM-PBSA energies can be employed (they are more stable and give a more accurate account of solvation effects than the raw QM/MM energies).
12. If you want to interpret a special state for which experimental X-ray, neutron, NMR, or EXAFS structural data are available, the combination of these methods and QM/MM geometry optimization is extremely powerful, as described in Section 2.6. Optimize all tentative structures and select the one that fits the experimental raw data best and also gives a small deviation in energy and structure from the vacuum-optimized structure.
13. Avoid studying reactions that change the net charge of the studied system. It is typically better to study internal electron or proton transfer, as described in Section 3.3. Further improvement can be obtained by neutralizing solvent-exposed residues. Still, such reactions give a lower accuracy than charge-conserving reactions.

By this approach, you solve or minimize the three most important problems with QM/MM calculations, the junctions (move them as far as possible from the reactive center), the size of the QM system (employ the big-QM approach), and the local-minima problem (calculate free energies with QTCP). Throughout this chapter, I have emphasized the importance of ensuring that the calculations are converged. The fact that some experimental data are reproduced does not guarantee that the calculations are reliable. Instead, you need to run several test calculations to ensure that the results can be trusted, as described in Section 3. Undoubtedly, QM/MM calculations are more complicated and error-prone than QM-cluster calculations. However, when carefully performed and tested, they can give accurate and reliable results for reactions in protein. In particular, they can be employed to understand the catalytic effect of the protein.

ACKNOWLEDGMENTS

This investigation has been supported by grants from the Swedish research council (Project 2014-5540) and from Knut and Alice Wallenberg Foundation (KAW 2013.0022). The computations were performed on computer resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lunarc at Lund University and HPC2N at Umeå University.

REFERENCES

André, I., Kesvatera, T., Jönsson, B., Åkerfeldt, K. S., & Linse, S. (2004). The role of electrostatic interactions in calmodulin-peptide complex formation. *Biophysical Journal*, *87*, 1929–1938.

Bashford, D., & Case, D. A. (2000). Generalized Born models of macromolecular solvation effects. *Annual Review of Physical Chemistry*, *51*, 129–152.

Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, *38*, 3098–3100.

Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, *98*, 5648–5652.

Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, *22*, 245–268.

Blomberg, M. R. A., Borowski, T., Himo, F., Liao, R.-Z., & Siegbahn, P. E. M. (2014). Quantum chemical studies of mechanisms for metalloenzymes. *Chemical Reviews*, *114*, 3601–3658.

Blumberger, J. (2010). Free energies for biological electron transfer from QM/MM calculation: Method, application and critical assessment. *Physical Chemistry Chemical Physics*, *10*, 5651–5667.

Caldararu, O., Andrejic, M., Cioloboc, D., Mata, R. A., & Ryde, U. (2016). QM/MM study of the reaction mechanism of sulfite oxidase. *Journal of Biological Inorganic Chemistry*. submitted.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G., & Skelton, N. J. (1996). *Protein NMR spectroscopy. Principles and practice*. London: Academic Press.

Chen, J., Liu, X., VandeVondele, J., Sulpizi, M., & Sprik, M. (2014). Redox potentials and acidity constants from density functional theory based molecular dynamics. *Accounts of Chemical Research*, *47*, 3522–3529.

Cobb, N., Conrads, T., & Hille, R. (2005). Mechanistic studies of Rhodobacter sphaeroides Me_2SO reductase. *The Journal of Biological Chemistry*, *280*, 11007.

Cooper, A. M., & Kästner, J. (2014). Averaging techniques for reaction barriers in QM/MM simulations. *Chemphyschem*, *15*, 3264–3269.

Cossi, M., Tomasi, J., & Cammi, R. (1995). Analytical expressions of the free energy derivatives for molecules in solution. Application to the geometry optimization. *International Journal of Quantum Chemistry*, *56*, 695–702.

Cramer, C. J. (2006). *Essentials of computational chemistry: Theories and models*. Chichester: J. Wiley & Sons.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*, 10089–10092.

Delcey, M. G., Pierloot, K., Phung, Q. M., Vancoillie, S., Lindh, R., & Ryde, U. (2014). Accurate calculations of geometries and singlet–triplet energy differences for active-site models of [NiFe] hydrogenase. *Physical Chemistry Chemical Physics*, *16*, 7927–7938.

Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., & Stewart, J. J. P. (1985). AM1: A new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, *107*, 3902–3909.

Dölker, N., Maseras, F., & Siegbahn, P. E. M. (2004). Stabilization of the adenosyl radical in coenzyme B 12—A theoretical study. *Chemical Physics Letters*, *386*, 174–178.
Dong, G., & Ryde, U. (2016). Protonation states of intermediates in the reaction mechanism of [NiFe] hydrogenase studied by computational methods. *Journal of Biological Inorganic Chemistry*, *21*, 2016, 383–394.
Ferré, N., Assfeld, X., & Rivail, J. L. (2002). Specific force field determination for the hybrid ab initio QM/MM LSCF method. *Journal of Computational Chemistry*, *23*, 610–624.
Flaig, D., Beer, M., & Ochsenfeld, C. J. (2012). Convergence of electronic structure with the size of the QM region: Example of QM/MM NMR Shieldings. *Journal of Chemical Theory and Computation*, *8*, 2260–2271.
Fock, V. A. (1930). An approximate method for quantum mechanical calculations. *Zeitsch. Phys.*, *62*, 795.
Freindorf, M., Shao, Y., Furlani, T. R., & Kong, J. (2005). Lennard–Jones parameters for the combined QM/MM method using the B3LYP/6-31G*/AMBER potential. *Journal of Computational Chemistry*, *26*, 1270–1278.
Gao, J., Amara, P., Alhambra, C., & Field, M. J. (1998). A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *The Journal of Physical Chemistry. A*, *102*, 4714–4721.
Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*, 449–461.
Götz, A. W., Clark, M. A., & Walker, R. C. (2014). An extensible interface for QM/MM molecular dynamics simulations with AMBER. *Journal of Computational Chemistry*, *35*, 95–108.
Gräter, F., Schwarzl, S. M., Dejaegere, A., Fischer, S., & Smith, J. C. (2005). Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *The Journal of Physical Chemistry. B*, *109*, 10474–10483.
Grimme, S. (2006). Semiempirical GGA-type density-functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, *27*, 1787–1799.
Grimme, S. (2011). Density functional theory with London dispersion corrections. *WIREs Computational Molecular Science*, *1*, 211–228.
Grimme, S., Antony, J., Ehrlich, S., & Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *The Journal of Chemical Physics*, *132*, 154104. 19 pages.
Grimme, S., Ehrlich, S., & Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry*, *32*, 1456–1465.
Hadt, R. G., Sun, N., Marshall, N. M., Hodgson, K. O., Hedman, B., Lu, Y., et al. (2012). Spectroscopic and DFT studies of second-sphere variants of the Type 1 copper site in Azurin: Covalent and nonlocal electrostatic contributions to reduction potentials. *Journal of the American Chemical Society*, *134*, 16701–16716.
Hansen, N., & van Gunsteren, W. E. (2014). Practical aspects of free-energy calculations: A review. *Journal of Chemical Theory and Computation*, *10*, 2632–2647.
Hartree, D. R. (1928). The wavemechanics of an atom with a non-Coulomb central field. I. Theory and methods. *Proceedings of the Cambridge Philological Society*, *24*, 89.
Hay, B. P., & Finke, R. G. J. (1986). Thermolysis of the cobalt-carbon bond of adenosylcobalamin. 2. Products, kinetics, and cobalt-carbon bond dissociation energy in aqueous solution. *American Chemical Society*, *108*, 4820–4829.
Hedegård, E. D., Kongsted, J., & Ryde, U. (2015). Multiscale modelling of the active site of [Fe]-hydrogenase: The H_2 binding site in open and closed protein conformations. *Angewandte Chemie*, *54*, 6246–6250.
Heimdal, J., Kaukonen, M., Srnec, M., Rulíšek, L., & Ryde, U. (2011). Reduction potentials and acidity constants of Mn superoxide dismutase calculated by QM/MM free-energy methods. *ChemPhysChem*, *12*, 3337–3347.

Heimdal, J., Rydberg, P., & Ryde, U. (2008). Protonation of the proximal histidine ligand in haem peroxidases. *The Journal of Physical Chemistry. B*, *112*, 2501–2510.
Heimdal, J., & Ryde, U. (2012). Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations. *Physical Chemistry Chemical Physics*, *14*, 12592–12604.
Hersleth, H.-P., Ryde, U., Rydberg, P., Görbitz, C. H., & Andersson, K. K. (2006). Structures of the high-valent metal-ion haem–oxygen intermediates in peroxidases, oxygenases and catalases. *Journal of Inorganic Biochemistry*, *100*, 460–476.
Hille, R., Hall, J., & Basu, P. (2014). The mononuclear molybdenum enzymes. *Chemical Reviews*, *114*, 3963–4038.
Himo, F., & Siegbahn, P. E. (2009). Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *Journal of Biological Inorganic Chemistry*, *14*, 643–651.
Hohenberg, P., & Kohn, W. (1964). Inhomogeneous electron gas. *Physical Review*, *136*(3B), B864–B871.
Honig, B., Sharp, K., & Yang, A.-S. (1997). Macroscopic models of aqueous solutions: Biological and chemical applications. *The Journal of Physical Chemistry*, *97*, 1101–1109.
Hopmann, K. H., & Himo, F. J. (2008). Quantum chemical modeling of the dehalogenation reaction of haloalcohol dehalogenase. *Journal of Chemical Theory and Computation*, *4*, 1129–1137.
Hsiao, Y.-W., Drakenberg, T., & Ryde, U. (2005). NMR structure determination of proteins supplemented by quantum chemical calculations: Detailed structure of the Ca^{2+} sites in the EGF34 fragment of protein S. *Journal of Biomolecular NMR*, *31*, 97–114.
Hsiao, Y.-W., & Ryde, U. (2006). Interpretation of EXAFS spectra for sitting-atop complexes with the help of computational methods. *Inorganica Chimica Acta*, *359*, 1081–1092.
Hsiao, Y.-W., Tao, Y., Shokes, J. E., Scott, R. A., & Ryde, U. (2006). EXAFS structure refinement supplemented by computational chemistry. *Physical Review B*, *74*, 214101.
Hu, L., Eliasson, J., Heimdal, J., & Ryde, U. (2009). Do quantum mechanical energies calculated for small models of protein active sites converge? *The Journal of Physical Chemistry. A*, *113*, 11793–11800.
Hu, L., Farrokhnia, M., Heimdal, J., Shleev, S., Rulíšek, L., & Ryde, U. (2011). Reorganisation energy for internal electron transfer in multicopper oxidases. *The Journal of Physical Chemistry. B*, *115*, 13111–13126.
Hu, L., Söderhjelm, P., & Ryde, U. (2011). On the convergence of QM/MM energies. *Journal of Chemical Theory and Computation*, *7*, 761–777.
Hu, L., Söderhjelm, P., & Ryde, U. (2013). Accurate reaction energies in proteins obtained by combining QM/MM and large QM calculations. *Journal of Chemical Theory and Computation*, *9*, 640–649.
Hu, H., & Yang, W. (2008). Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annual Review of Physical Chemistry*, *59*, 573–601.
Iftimie, R., Salahub, D., Wei, D., & Schofield, J. (2000). Using a classical potential as an efficient importance function for sampling from an *ab initio* potential. *Journal of Chemical Physics*, *113*, 4852–4862.
Jensen, F. (2007). *Introduction to computational chemistry*. Chichester: J. Wiley & Sons.
Jensen, K. P., & Ryde, U. (2003). Theoretical prediction of the Co-C bond strength in cobalamins. *The Journal of Physical Chemistry. A*, *107*, 7539–7545.
Jensen, K. P., & Ryde, U. (2005). How the Co-C bond is cleaved in coenzyme B12 enzymes, a theoretical study. *Journal of the American Chemical Society*, *127*, 9117–9128.
Jensen, K. P., & Ryde, U. (2009). Cobalamins uncovered by modern electronic structure calculations. *Coordination Chemistry Reviews*, *253*, 769–778.

Kaukonen, M., Söderhjelm, P., Heimdal, J., & Ryde, U. (2008a). Proton transfer at metal sites in proteins studied by quantum mechanical free-energy perturbations. *Journal of Chemical Theory and Computation*, *4*, 985–1001.

Kaukonen, M., Söderhjelm, P., Heimdal, J., & Ryde, U. (2008b). A QM/MM-PBSA method to estimate free energies for reactions in proteins. *The Journal of Physical Chemistry. B*, *112*, 12537–12548.

Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, *3*, 300–313.

Klamt, A., & Schüürmann, G. (1994). COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions*, *2*, 799–805.

Kleywegt, G. J., & Jones, T. A. (1997). Model building and refinement practice. *Methods in Enzymology*, *227*, 208.

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. (2000). Calculating structures and free energies of complex molecules: Combining Molecular mechanics and continuum models. *Accounts of Chemical Research*, *33*, 889–897.

König, G., & Boresch, S. (2011). Non-Boltzmann sampling and Bennett's acceptance ratio method: How to profit from bending the rules. *Journal of Computational Chemistry*, *32*, 1082–1090.

König, P. H., Hoffman, M., Trauenheim, T., & Cui, Q. J. (2005). A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. *Physical Chemistry B*, *109*, 9082–9095.

Kozlowski, P. M., Kumar, M., Piecuch, P., Li, W., Bauman, N. P., Hansen, J. A., et al. (2012). The cobalt−methyl bond dissociation in methylcobalamin: New benchmark analysis based on density functional theory and completely renormalized coupled-cluster calculations. *Journal of Chemical Theory and Computation*, *8*, 1870–1894.

Kwiecien, R. A., Khavrutskii, I. V., Musaev, D. G., Morokuma, K., Banerjee, R., & Paneth, P. (2006). Computational insights into the mechanism of radical generation in B_{12}-dependent methylmalonyl-CoA mutase. *Journal of the American Chemical Society*, *128*, 1287.

Lee, C. T., Yang, W. T., & Parr, R. G. (1988). Development of the Colic-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, *37*, 785–789.

Li, J., Andrejic, M., Mata, R. A., & Ryde, U. (2015). A computational comparison of oxygen atom transfer catalyzed by DMSO reductase with Mo and W. *European Journal of Inorganic Chemistry*, *2015*, 3580–3589.

Li, J., Farrokhnia, M., Rulíšek, L., & Ryde, U. (2015). Catalytic cycle of multicopper oxidases studied by combined quantum- and molecular-mechanical free-energy perturbation methods. *The Journal of Physical Chemistry. B*, *119*, 8268–8284.

Li, J.-L., Mata, R. A., & Ryde, U. (2013). Large density-functional and basis-set effects for the DMSO reductase catalyzed oxo-transfer reaction. *Journal of Chemical Theory and Computation*, *9*, 1799–1807.

Li, X., Siegbahn, P. E. M., & Ryde, U. (2015). A simulation of the isotropic EXAFS spectra for the S2 and S3 structures of the oxygen evolving complex in photosystem II. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 3979–3984.

Li, X., Sproviero, E. M., Ryde, U., Batista, V. S., & Chen, G. (2013). Theoretical EXAFS studies of a model of the oxygen-evolving complex of photosystem II obtained with the quantum cluster approach. *International Journal of Quantum Chemistry*, *113*, 474–478.

Liao, R.-Z., & Thiel, W. (2013). Convergence in the QM-Only and QM/MM modeling of enzymatic reactions: A case study for acetylene hydratase. *Journal of Combinatorial Chemistry*, *34*, 2389–2397.

Liao, R.-Z., Yu, G., & Himo, F. J. (2011). Quantum chemical modeling of enzymatic reactions: The case of decarboxylation. *Journal of Chemical Theory and Computation*, *7*, 1494–1501.

Lin, H., & Truhlar, D. G. J. (2005). Redistributed charge and dipole schemes for combined quantum mechanical and molecular mechanical calculations. *Physical Chemistry A*, *109*, 3991–4004.

Lin, H., & Truhlar, D. G. (2007). QM/MM: What have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts*, *117*, 185–199.

Liu, J., Chakraborty, S., Hosseinzadeh, P., Yu, Y., Tian, S., Petrik, I., et al. (2014). Metalloproteins containing cytochrome, iron–sulfur, or copper redox centers. *Chemical Reviews*, *114*, 4366–4469.

Lonsdale, R., Harvey, J. N., & Mulholland, A. J. (2010). Compound I reactivity defines alkene oxidation selectivity in cytochrome P450cam. *The Journal of Physical Chemistry. B*, *114*, 1156–1162.

Luzhkov, V., & Warshel, A. (1992). Microscopic models for quantum mechanical calculations of chemical processes in solutions: LDIAMPAC and SCAAS/AMPAC calculations of solvation energies. *Journal of Computational Chemistry*, *13*, 199–213.

Mackerell, A. D. (2004). Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry*, *25*, 1584–1604.

Manzoni, F., Oksanen, E., Logan, D., & Ryde, U. (2016). Structural of galectin 3 by a combination of neutron and X-ray crystallography and quantum mechanics. *Journal of Chemical Theory and Computation*. submitted.

Marsh, E. N. G. (2000). Coenzyme-B_{12}-dependent glutamate mutase. *Bioorganic Chemistry*, *28*, 176–189.

Martin, B. D., & Finke, R. G. (1992). Methylcobalamin's full- vs. half-strength cobalt-carbon sigma bonds and bond dissociation enthalpies: $A>10^{15}$ Co-CH_3 homolysis rate enhancement following one-antibonding-electron reduction of methlycobalamin. *Journal of the American Chemical Society*, *114*, 585.

Metz, S., & Thiel, W. (2009). A combined QM/MM study on the reductive half-reaction of xanthine oxidase: Substrate orientation and mechanism. *Journal of the American Chemical Society*, *2009*(131), 14885–14902.

Mikulskis, P., Genheden, S., & Ryde, U. (2014). Large-scale test of free-energy simulation estimates of protein-ligand binding affinities. *Journal of Chemical Information and Modeling*, *54*, 2794–2806.

Møller, C., & Plesset, M. S. (1934). Note on an approximation treatment for many-electron systems. *Physics Review*, *46*, 618.

Murphy, R. B., Philipp, D. M., & Friesner, R. A. (2000). A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *Journal of Computational Chemistry*, *21*, 1442–1457.

Nicoll, R. M., Hindle, S. A., MacKenzie, G., Hillier, I. H., & Burton, N. A. (2001). Quantum mechanical/molecular mechanical methods and the study of kinetic isotope effects: Modelling the covalent junction region and application to the enzyme xylose isomerase. *Theoretical Chemistry Accounts*, *106*, 105–112.

Nilsson, K., & Ryde, U. (2004). Protonation status of protein ligands can be determined by quantum refinement. *Journal of Inorganic Biochemistry*, *98*, 1539–1546.

Olsson, M. A., Söderhjelm, P., & Ryde, U. (2016). Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level. *Journal of Computational Chemistry*, *37*, 1589–1600.

Olssson, M. H. A., Hong, G., & Warshel, A. (2003). Frozen density functional free energy simulations of redox proteins: Computational studies of the reduction potential of plastocyanin and rusticyanin. *Journal of the American Chemical Society*, *125*, 5025–5039.

Padmakumar, R., Padmakumar, R., & Banerjee, R. (1997). Evidence that cobalt–carbon bond homolysis is coupled to hydrogen atom abstraction from substrate in methylmalonyl-CoA mutase. *Biochemistry*, *36*, 3713–3718.
Perdew, J. P. (1986). Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B*, *33*, 8822–8824.
Perdew, J. P., Burke, K., & Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Physical Review Letters*, *77*, 3865–3868.
Philipp, D. M., & Friesner, R. A. (1999). *Journal of Computational Chemistry*, *20*, 1468–1494.
Plotnikov, N. V., Kamerlin, S. C. L., & Warshel, A. (2011). Paradynamics: An effective and reliable model for ab initio QM/MM free-energy calculations and related tasks. *The Journal of Physical Chemistry. B*, *115*, 7950–7962.
Poulsen, T. D., Kongsted, J., Osted, A., Ogilby, P. R., & Mikkelsen, K. V. (2001). The combined multiconfigurational self-consistent-field molecular mechanics wave function approach. *The Journal of Chemical Physics*, *115*, 2393–2400.
Pu, J., Gao, J., & Truhlar, D. G. (2005). Generalized hybrid-orbital method for combining density functional theory with molecular mechanicals. *Chemphyschem*, *6*, 1853–1865.
Raghavachari, K., Tucks, G. W., Pople, J. A., & Head-Gordon, M. (1989). A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters*, *157*, 479–483.
Rappoport, D., & Furche, F. (2010). Property-optimized Gaussian basis sets for molecular response calculations. *The Journal of Chemical Physics*, *133*, 134105.
Reddy, M. R., & Erion, M. D. (2007). Relative binding affinities of fructose-1,6-bisphosphatase inhibitors calculated using a quantum mechanics-based free energy perturbation method. *Journal of the American Chemical Society*, *129*, 9296–9297.
Retegan, M., Milet, A., & Jamet, H. (2009). Exploring the binding of inhibitors derived from tetrabromobenzimidazole to the CK2 protein using a QM/MM-PB/SA approach. *Journal of Chemical Information and Modeling*, *49*, 963–971.
Riahi, S., & Rowley, C. N. (2014). The CHARMM–TURBOMOLE interface for efficient and accurate QM/MM molecular dynamics, free energies, and excited state properties. *Journal of Computational Chemistry*, *35*, 2076–2086.
Riccardi, D., Li, G., & Cui, Q. (2004). Importance of van der Waals interactions in QM/MM simulations. *The Journal of Physical Chemistry. B*, *108*, 6467–6478.
Rod, T. H., & Ryde, U. (2005a). Quantum mechanical free energy barrier for an enzymatic reaction. *Physical Review Letters*, *94*, 138302. 4 pages.
Rod, T. H., & Ryde, U. (2005b). Free energy barriers at the density functional theory level: Methyl transfer catalyzed by catechol O-methyltransferase. *Journal of Chemical Theory and Computation*, *1*, 1240–1251.
Román-Meléndez, G. D., von Glehn, P., Harvey, J. N., Mulholland, A. J., & Marsh, E. N. G. (2014). Role of active site residues in promoting cobalt–carbon bond homolysis in adenosylcobalamin-dependent mutases revealed through experiment and computation. *Biochemistry*, *53*, 169–177.
Rulíšek, L., & Ryde, U. (2006). Structure of reduced and oxidised manganese superoxide dismutase—A combined computational and experimental approach. *The Journal of Physical Chemistry. B*, *110*, 11511–11518.
Ryde, U. (1995). On the role of Glu68 in alcohol dehydrogenase. *Protein Science*, *4*, 1124–1132.
Ryde, U. (1996). The coordination of the catalytic zinc ion in alcohol dehydrogenase studied by combined quantum chemical and molecular mechanical calculations. *Journal of Computer-Aided Molecular Design*, *10*, 153–164.
Ryde, U. (2003). Combined quantum and molecular mechanics calculations on metalloproteins. *Current Opinion in Chemical Biology*, *7*, 136–142.

Ryde, U. (2007). Accurate metal-site structures in proteins obtained by combining experimental data and quantum chemistry. *Dalton Transactions*, *2007*, 607–625.
Ryde, U., Dong, G., Li, J., Andrejic, M., & Mata, R. A. (2016). Computational studies of molybdenum and tungsten enzymes. In R. Hille, M. Kirk, & C. Schulzke (Eds.), *Molybdenum and tungsten enzymes*. in press.
Ryde, U., Hsiao, Y.-W., Rulíšek, L., & Solomon, E. I. (2007). Identification of the peroxy adduct in multicopper oxidases by a combination of computational chemistry and extended X-ray absorption fine-structure measurements. *Journal of the American Chemical Society*, *129*, 726–727.
Ryde, U., Mata, R. A., & Grimme, S. (2011). Does DFT-D estimate accurate energies for the binding of ligands to metal complexes? *Dalton Transactions*, *40*, 11176–11183.
Ryde, U., & Nilsson, K. (2003). Quantum chemistry can improve protein crystal structures locally. *Journal of the American Chemical Society*, *125*, 14232–14233.
Ryde, U., Olsen, L., & Nilsson, K. (2002). Quantum chemical geometry optimisations in proteins using crystallographic raw data. *Journal of Combinatorial Chemistry*, *23*, 1058–1070.
Ryde, U., & Olsson, M. H. M. (2001). Structure, strain, and reorganization energy of blue copper models in the protein. *International Journal of Quantum Chemistry*, *81*, 335–347.
Ryde, U., Olsson, M. H. M., & Pierloot, K. (2001). The structure and function of blue copper proteins. In L. A. Eriksson (Ed.), *Theoretical and computational chemistry*: *Vol. 9. Theoretical biochemistry. Processes and properties of biological systems* (pp. 1–56). Amsterdam: Elsevier.
Ryde, U., & Söderhjelm, P. (2016). Ligand-binding affinity estimates supported by quantum-mechanical methods. *Chemical Reviews*, *116*, 5520–5566.
Senn, H. M., & Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angewandte Chemie, International Edition*, *48*, 1198–1229.
Senn, H. M., Thiel, S., & Thiel, W. (2005). Enzymatic hydroxylation in p-hydroxybenzoate hydroxylase: A case study for QM/MM molecular dynamics. *Journal of Chemical Theory and Computation*, *1*, 494–505.
Sevastik, R., & Himo, F. (2007). Quantum chemical modeling of enzymatic reactions: The case of 4-oxalocrotonate tautomerase. *Bioorganic Chemistry*, *35*, 444–457.
Sharma, P. K., Chu, Z. T., Olsson, M. H., & Warshel, A. (2007). A new paradigm for electrostatic catalysis of radical reactions in vitamin B_{12} enzymes. *Proceedings of the National academy of Sciences of the United States of America*, *104*, 9661.
Sherwood, P., de Vries, A. H., Collins, S. J., Greatbanks, S. P., Burton, N. A., Vincent, M. A., et al. (1997). Computer simulation of zeolite structure and reactivity using embedded cluster methods. *Faraday Discussions*, *106*, 79–92.
Sherwood, P., de Vries, A. H., Guest, M. F., Schreckenbach, G., Catlow, C. R. A., French, S. A., et al. (2003). QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *Journal of Molecular Structure*, *632*, 1–28.
Siegbahn, P. E. M., Blomberg, M. R. A., & Chen, S.-L. (2010). Significant van der Waals effects in transition metal complexes. *Journal of Chemical Theory and Computation*, *6*, 2040–2044.
Singh, U. C., & Kollman, P. A. (1986). A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH_3Cl + Cl exchange reaction and gas phase protonation of polyethers. *Journal of Combinatorial Chemistry*, 7, 718–730.
Söderhjelm, P., Husberg, C., Strambi, A., Olivucci, M., & Ryde, U. (2009). Protein influence on electronic spectra modelled by multipoles and polarisabilities. *Journal of Chemical Theory and Computation*, *5*, 649–658.

Söderhjelm, P., & Ryde, U. (2006). Combined computational and crystallographic study of the oxidised states of [NiFe] hydrogenase. *Journal of Molecular Structure (THEOCHEM)*, *770*, 199–219.

Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2012). Computational enzymatic catalysis—Clarifying enzymatic mechanisms with the help of computers. *Physical Chemistry Chemical Physics*, *14*, 12431–12441.

Stewart, J. J. P. (1989). Optimization of parameters for semiempirical methods. I: Method. *Journal of Combinatorial Chemistry*, *10*, 209–220.

Stewart, J. J. P. (2007). Optimization of parameters for semiempirical methods. V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling*, *13*, 1173–1213.

Sumner, S., Söderhjelm, P., & Ryde, U. (2013). Effect of geometry optimisations on QM-cluster and QM/MM studies of reaction energies in proteins. *Journal of Chemical Theory and Computation*, *9*, 4205–4214.

Sumowski, C. V., & Ochsenfeld, C. J. (2009). A convergence study of QM/MM isomerization energies with the selected size of the QM region for peptidic systems. *Physical Chemistry A*, *113*, 11734–11741.

Sure, R., & Grimme, S. (2015). Comprehensive benchmark of association (free) energies of realistic host−guest complexes. *Journal of Chemical Theory and Computation*, *11*, 3785–3801.

Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., & Morokuma, K. (1996). ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)$_3$)$_2$ + H2 oxidative addition. *The Journal of Physical Chemistry*, *100*, 19357–19363.

Tao, J., Perdew, J. P., Staroverov, V. N., & Scuseria, G. E. (2003). Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Physical Review Letters*, *91*, 146401. 4 pages.

Théry, V., Rinaldi, D., Rivail, J. L., Maigret, B., & Ferenczy, G. J. (1994). Quantum mechanical computations on very large molecular systems: The local self-consistent field method. *Computers & Chemistry*, *15*, 269–282.

Thiel, W. (2014). Semiempirical quantum–chemical methods. *WIREs Computational Molecular Science*, *4*, 145–157.

Tian, B., Strid, Å., & Eriksson, L. A. (2011). Catalytic roles of active-site residues in 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase: An ONIOM/DFT study. *The Journal of Physical Chemistry. B*, *115*, 1918–1926.

Tomasi, J., Mennucci, B., & Cammi, R. (2005). Quantum mechanical continuum solvation models. *Chemical Reviews*, *105*, 2999–3094.

Van den Bosch, M., Swart, M., Snijders, J. G., Berendsen, H. J. C., Mark, A. E., Oostenbrink, C., et al. (2005). Calculation of the redox potential of the protein azurin and some mutants. *Chembiochem*, *6*, 738–746.

VandeVondele, J., & Rothlisberger, U. (2002). Canonical adiabatic free energy sampling (CAFES): A novel method for the exploration of free energy surfaces. *The Journal of Physical Chemistry. B*, *106*, 203–208.

Van Severen, M.-C., Andrejic, M., Li, J.-L., Starke, K., Mata, R. A., Nordlander, E., et al. (2014). A quantum-mechanical study of the reaction mechanism of sulfite oxidase. *Journal of Biological Inorganic Chemistry*, *19*, 1165–1179.

Von Lilienfeld, O. A., Tavernelli, I., & Rothlisberger, U. (2005). Variational optimization of effective atom centered potentials for molecular properties. *The Journal of Chemical Physics*, *122*, 014113. 6 pages.

Vreven, T., Byun, K. S., Komáromi, I., Dapprich, S., Montgomery, J. A., Morokuma, K., et al. (2006). Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *Journal of Chemical Theory and Computation*, *2*, 815–826.

Warshel, A., & Levitt, M. (1976). Theoretical studies of enzymic reactions: Electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, *103*, 227–249.

Weigend, F., & Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*, *7*, 3297–3305.

Werner, H.-J., & Schütz, M. (2011). An efficient local coupled cluster method for accurate thermochemistry of large systems. *The Journal of Chemical Physics*, *135*, 144116. 15 pages.

Wood, R. H., Yezdimer, E. M., Sakane, S., Barriocanal, J. A., & Doren, D. J. (1999). Free energies of solvation with quantum mechanical interaction energies from classical mechanical simulations. *Journal of Chemical Physics*, *110*, 1329–1337.

Woods, C. J., Shaw, K. E., & Mulholland, A. J. (2015). Combined quantum mechanics/molecular mechanics (QM/MM) simulations for protein−ligand complexes: Free energies of binding of water molecules in influenza neuraminidase. *The Journal of Physical Chemistry. B*, *119*, 997–1001.

Zhang, Y. (2006). Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theoretical Chemistry Accounts*, *116*, 43–50.

Zhang, Y., Liu, H., & Yang, W. J. (2000). Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *Chemical Physics*, *112*, 3483–3492.

Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics*, *22*, 1420–1427.