

Modelling the reactivity of glutamate mutase and heme dioxygenase enzymes



Patrick von Glehn

A thesis submitted to the University of Bristol in accordance with the
requirements of the Degree of Doctor of Philosophy in the School of Chemistry,
Faculty of Science.

September 2014

52,243 words

Abstract

Adenosylcobalamin (AdoCbl) serves as a reservoir for the 5'-deoxyadenosyl radical, which is generated in enzyme by the homolytic cleavage of a Co-C bond and harnessed to initiate radical reactions by abstracting a hydrogen from the substrate. How these enzymes increase the rate of Co-C bond cleavage by an estimated 12 orders of magnitude, whether the 5'-deoxyadenosyl radical exists as a metastable or transient intermediate and how the first steps of the reaction are coupled are key unresolved questions.

The Co-C bond breaking and hydrogen abstraction steps were modelled in AdoCbl dependent glutamate mutase with MD simulations, adiabatic mapping and umbrella sampling simulations using a novel empirical valence bond (EVB) potential, which was calibrated to high level *ab initio* and DFT calculations. This potential was found to compare favourably with the results of QM/MM calculations. Hydrogen bonding with the protein stabilises the dissociated 5'-deoxyadenosyl radical and induce conformational change, guiding the C5' radical centre towards the substrate hydrogen to be abstracted.

The heme dioxygenase enzymes Indoleamine 2,3-dioxygenase (IDO) and tryptophan 2,3-dioxygenase (TDO) catalyse the first step in the metabolism of L-tryptophan (L-Trp) by insertion of both atoms of heme-bound O₂ into the substrate. In an attempt to improve understanding of the differences in substrate binding and reactivity between these enzymes, molecular dynamics (MD) simulations, MM/PBSA binding free energy calculations and reaction modelling with hybrid quantum mechanics/molecular mechanics (QM/MM) adiabatic mapping calculations were performed. Starting with crystal structures for a bacterial TDO (XcTDO) and human IDO (hIDO), reactivity and binding of IDO, TDO and the H55A mutant TDO with L-Trp, D-tryptophan (D-Trp) and 1-methyl-L-tryptophan (1-Me-L-Trp) were investigated.

Differences in experimental K_{MS}s were partially rationalised by analysis of substrate-protein interactions and calculated binding free energies. Although the calculated barriers were unable to rank correctly the active systems, they were able to predict whether a particular system was active, slightly active or inactive. Differences in reactivity were related to the varying ability of the systems to position optimally the substrate in relation to the heme-bound O₂.

This thesis is dedicated to my trusted friend and esteemed colleague,
Michael Alexander Lewis Limb,
who has been with me every step of the way.

Acknowledgements

I would like to express my most heartfelt thanks to all of my friends and colleagues in the CCC. Whilst my relationship with my terminal has not always been a joyous one, I can honestly say that working amongst such warm and unassuming people for the past four years has been an absolute pleasure. Particular thanks goes to my fellow Super Friends for their solidarity throughout and to Laura Rush for her efforts in nurturing the CCC social scene.

Warm thanks of course goes to Jeremy Harvey and Adrian Mulholland, whose complementary supervising styles have led to a productive collaboration, the fruits of which lie within these pages. I am grateful to the following people, who have helped me with the work described in this thesis: Scott Habershon, Marc van der Kamp, David Tew, Wataru Mizukami, Judith Rommel and Christopher Woods. Thanks to Neil Marsh and Gabriel Román-Meléndez for a productive collaboration and two enjoyable trips to the States. Two people in particular have spent countless hours patiently transferring their knowledge to me. Special thanks goes to Richard Lonsdale for helping me find my feet in the bewildering world of computational chemistry and to David Glowacki for his mentorship in recent years.

Finally, I'd like to express my enormous gratitude for the constant love and support of my parents, brother and Bristol family.

Author's declaration

I declare that the work in this thesis was carried out in accordance with the regulations of the University of Bristol and no part of the thesis has been submitted for any other academic award. The work is original, except where indicated by special reference in the text and any work done in collaboration with, or with the assistance of others, is indicated as such. Any views expressed in the thesis are those of the author.

Patrick von Glehn

University of Bristol

September 2014

Table of Contents

Chapter 1. Modelling enzyme catalysed reactions	1
1.1. Potential energy surfaces	2
1.2. Calculating physical properties.....	9
Chapter 2. Glutamate mutase.....	14
2.1. Introduction.....	14
2.2. Review of previous computational studies	17
Chapter 3. Glutamate mutase PES.....	20
3.1. Quantum mechanical PES	20
3.2. Molecular mechanics PES	23
3.3. QM/MM PES.....	26
3.4. Empirical valence bond (EVB) PES.....	27
Chapter 4. Characterising the Co-C associated and dissociated states.....	33
4.1. Introduction.....	33
4.2. Computational details	35
4.3. Results and discussion	37
Chapter 5. Modelling Co-C bond breaking and hydrogen abstraction.....	60
5.1. Introduction.....	60
5.2. Computational details	61
5.3. Results and discussion	66
Chapter 6. Glutamate mutase conclusions.....	87
6.1. Conclusions.....	87
6.2. Further work	93
Chapter 7. Introduction to heme dioxygenase enzymes	96
7.1. Introduction.....	96
7.2. Reaction mechanism.....	101
Chapter 8. MD study of substrate binding in IDO and TDO.....	105
8.1. Introduction.....	105
8.2. Computational details	106
8.3. Results and discussion	112

Chapter 9. QM/MM study of reactivity in IDO and TDO.....	129
9.1. Introduction.....	129
9.2. Computational details	129
9.3. Results and discussion	135
Chapter 10. Heme dioxygenase conclusions	159
10.1. Conclusions.....	159
10.2. Further work	164

Chapter 1. Modelling enzyme catalysed reactions

In terms of the themes explored and questions asked, the research that we as computational chemists conduct is fundamentally no different from that of our experimental counterparts. The tools that we use for our research are physics based mathematical models. The implementation of these models in an ever-increasing body of academic and commercial software packages has allowed their widespread exploitation in the pursuit of greater understanding of the behaviour of matter on the molecular level.

The work described in this thesis is concerned with the structure, dynamics and reactivity of enzymes. The approaches used here rely on two approximations of the laws of physics. The first is the Born-Oppenheimer approximation, which allows for the separation of the motion of nuclei and electrons.¹ In this regime, the potential energy of a system is a direct function of the nuclear coordinates. The multidimensional surface relating the positions in space of all nuclei in a system to the potential energy is known as the potential energy surface (PES). The second approximation is that the motion of the nuclei on this PES can be described using Newton's classical equations of motion.²

Although we know that matter behaves according to the laws of quantum dynamics, classical dynamics has proven to be a surprisingly good approximation for the calculation of many molecular properties.³ Whilst the Born-Oppenheimer approximation is almost always invoked in computational chemistry studies, there are many phenomena that cannot be accurately modelled with a purely classical description of the dynamics. Proper treatment of nuclear quantum effects is required for accurate modelling of many time-dependent properties such as the finer details of the dynamics of chemical reactions or quantum tunnelling rates.^{4,5} Even some equilibrium properties require proper treatment of nuclear quantum effects, such as zero-point energy and phase transitions in water.^{6,7} Whilst all calculations described here have so far employed the classical dynamics approximation, a natural extension to some of this work would be the incorporation of nuclear quantum effects such as a proper modelling of hydrogen tunnelling. Indeed,

although not reported here, some preliminary work by the author and collaborators has already begun in this direction.

Most of the methodological challenges faced in the conduct of this work can be broken down into two types. The first is how to calculate a sufficiently accurate model of the PES at a computational expense low enough to enable adequate exploration of the configuration space. The second type of challenge is how to explore efficiently this configuration space in order to calculate physically meaningful properties from the models.

Many of the methods used in this work (e.g. density functional theory or molecular dynamics simulations) have become very standard tools in the field and will be familiar to any computational chemist conducting similar research. Where this is the case, the method will not be discussed in detail. For further background information, the reader is directed to a number of excellent introductory texts.^{1,8–10}

1.1. Potential energy surfaces

1.1.1. Quantum mechanics

In principle, the exact PES of a system can be obtained by solving the time-independent electronic Schrödinger equation. In practice this is impossible for all but the simplest of systems. A huge proportion of all research undertaken in the field of theoretical chemistry has been dedicated to developing methods for calculating the closest approximation to the exact solution to the electronic Schrödinger equation possible for a given system size and computational expense.⁸ One highly accurate method, often cited as the gold standard of electronic structure methods is CCSD(T).^{11,12} Historically, slow convergence towards the complete basis set limit, combined with poor scaling with respect to basis set size has confined the use of CCSD(T) to systems of the order of 10 heavy atoms. Large improvements in convergence with respect to basis set size have been obtained by the development of explicitly correlated F12 methods.^{13,14} CCSD(T)-F12 has been shown to give results of chemical accuracy with basis sets as small as valence-double- ζ , opening up these highly accurate methods to ever larger systems.¹⁵

Another very widely used quantum mechanics based method for calculating PES is Kohn-Sham density functional theory (DFT).¹⁶ The central premise of DFT is that the ground state electronic energy of a system is uniquely defined by the electron density.¹⁷ Unfortunately, the exact functional relating the electron density of a system to its ground state energy is unknown. Over the past 50 years, a vast number of approximate functionals have been developed and applied, each giving more or less accurate energies for a given type of system.¹⁸ DFT has become extremely popular due to its ability to calculate energies at considerably less computational expense than high level *ab initio* wavefunction methods, whilst often giving results of comparable accuracy. DFT is however, also capable of generating serious errors in certain cases, such as in calculation of the bond dissociation energies for cobalamins discussed later in this work.¹⁹ One of the causes of these errors is that standard DFT neglects to model dispersion forces.²⁰ Some groups have specifically parametrised functionals to account for this, such as the M06 family of functionals.²¹ Another approach (used in this work) is to add an empirical dispersion correction to the DFT energy, such as that developed in the group of Grimme.²²

Though cheaper than high level *ab initio* methods, standard DFT is still too expensive for application to large systems such as full enzymes in solution. Many groups choose to truncate their enzyme models to the order of 200 of the most essential active site atoms and use standard DFT.²³ Increasingly, research groups are turning to linear scaling DFT methods which are now able to model systems of tens of thousands of atoms.²⁴

1.1.2. Molecular mechanics

Still by far the most widely used method of calculating PES for protein systems is molecular mechanics. Molecular mechanics is a method whereby the energy of a system is related to the nuclear coordinates by numerous simple, additive, analytic functions. For example, all bonds in a system could be modelled as harmonic oscillators. In this case, each bond will have an energy term associated with it of the form shown in Equation 1.

$$E_{bond} = k_{ij}(R_{ij} - Req_{ij})^2$$

Equation 1 Energy term for a bond between atoms i and j from a typical molecular mechanics force field. k_{ij} is the force constant R_{ij} is the distance between atoms i and j. Req_{ij} is the equilibrium bond length. k_{ij} and Req_{ij} are parameters specific to the pair of atom types i and j.

Two parameters are associated with each type of bond: a force constant and an equilibrium bond length. The total bonded component of the total energy will be the sum of all individual bonding terms. Further functions can be added, for example to describe angle bending, dihedral angle rotation and non-bonded interactions. The non-bonded interactions tend to be pair wise functions, such as that given in Equation 2 for the van der Waals' components of the non-bonded interactions. The combination of all of these functions and their associated parameters is known as a force field. Two families of force fields used in this work and by many other research groups for simulation of proteins are the CHARMM and Amber force fields.^{25,26} Because of the mathematical simplicity and additive nature of force fields, the computational expense is intrinsically low and scales very well with system size in comparison to quantum mechanics based methods. It is now routine to run molecular dynamics simulation of systems containing hundreds of thousands of atoms for hundreds of nanoseconds in a wall-clock timescale of days.²⁷

Because the functional forms of molecular mechanics force fields are not rigorously grounded in physics, they must be carefully parametrised against data from experiment or quantum mechanics calculations. Much work has been put into rigorously parametrising force fields such as the CHARMM or Amber in order to reproduce accurately certain equilibrium properties of proteins, nucleic acids and a handful of other common types of biological molecules. Very often a researcher will want to model a molecule for which there are no parameters in the standard force fields. In these cases, unless there are some parameters available elsewhere in the literature, the researcher will be obliged to undergo the often painstaking, tedious and treacherous task of parametrisation.

$$E_{vdW} = \varepsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{R_{ij}} \right)^{12} - 2 \left(\frac{R_{min_{ij}}}{R_{ij}} \right)^6 \right]$$

Equation 2 Energy term modelling the van der Waals' interactions between atoms i and j. This is the form of the Lennard-Jones potential found in the CHARMM family of force fields.^{28,29}

1.1.3. QM/MM methods

Standard molecular mechanics force fields are typically only parametrised to model molecules close to their equilibrium structures. Therefore, on their own, standard

molecular mechanics methods cannot be used to model reactions. The combination of quantum mechanics and molecular mechanics (QM/MM) allows for accurate modelling (using quantum mechanics) of the region in which the reaction is taking place, whilst still accounting for the influence of the surrounding environment on the reaction (using molecular mechanics). Only the QM/MM methods used in this work will be described here. However, numerous other schemes do exist to couple molecular mechanics and quantum mechanics methods, many of which are discussed in the review cited here.³⁰

In this work we have used our own in-house QM/MM code, QoMMMa, which employs a QM/MM scheme described in brief here and in detail in the following reference.³¹ Each system is divided into two regions: the QM region and the MM region. The total energy of the system at a given geometry is given by the sum of the energies resulting from a QM calculation and an MM calculation. Where the QM/MM boundary is separated by covalent bonds, the link atom method is used, whereby capping hydrogen atoms are added to the QM calculation at positions along the relevant bonds.³² The QM calculation is performed on the QM region plus link atoms. The point charges associated with the MM atoms are included in the QM Hamiltonian. The point charges associated with the MM atoms on the QM/MM boundary are set to zero, as are the charges associated with some additional nearby MM atoms. This is in order to avoid unphysical effects due to interactions between those charges and the link atoms. Further MM charge redistribution is usually required so as not to introduce or remove charge from the system as a result of the charge zeroing and QM region assignment.

Bond, angle and dihedral terms are included in the MM calculation only if the term involves at least one MM atom. All charge-charge interactions between QM and MM atoms are ignored because these are already accounted for by including the MM point charges in the QM Hamiltonian. All van der Waals' interactions between QM and MM atoms are included in the MM calculation. Interactions involving only MM atoms are calculated as normal.

All of the systems studied here have been partitioned into QM regions of roughly 60 atoms (usually including one transition metal atom each) and MM regions of roughly 10000-20000 atoms. Using DFT for the QM calculations, the QM/MM scheme described here permits roughly one force evaluation every 20 minutes for these systems. This is acceptably fast for techniques of exploring the PES based on energy minimisation such as adiabatic mapping (described later). The number of force evaluations required for the calculation of a full adiabatic mapping pathway tends to be of the order of 1000,

corresponding to a wall-clock time of the order of a few weeks. In order to run molecular dynamics (MD) simulations long enough to be of any use, far more force evaluations are required. For example a 1 ns MD simulation with a 1 fs timestep would require one million force evaluations, which would take of the order of 50 years to simulate using this QM/MM scheme. Some groups do run QM/MM MD simulations of similar systems, for example using the Car-Parrinello method, although the length of simulation tends to be limited to tens of picoseconds.³³

1.1.4. Empirical valence bond methods (EVB)

It is possible to develop PES capable of modelling reactions that have functional forms cheap enough to perform the many force evaluations required for typical MD simulations. As with standard molecular mechanics force fields, these analytic PES need to be carefully fitted to some reference data. Simple functions, such as those shown in Equation 2 and Equation 3, can produce a good approximation of a system's PES in the low energy areas visited at equilibrium. In contrast, the higher energy areas of a PES, such as those involved in reactive processes, typically have much more complicated and irregular topologies. This not only makes fitting analytic functions to a particular high energy area of a PES challenging, but it also severely limits the transferability of these functions to different systems. The large overhead of researcher time required to accurately parametrise a PES for each new class of problem is a major drawback and one of the main why these methods are much less widely used than those discussed above. Nevertheless, the development and application of these types of PES have a long history and are still active areas of research. Different groups take a wide variety of approaches to this problem, a discussion of which is beyond the scope of this work.³⁴⁻³⁹

In this work, the empirical valence bond method (EVB) was used as described in brief here and in detail in the following reference.⁴⁰ This method is conceptually similar, but not identical to the EVB method used by the Warshel group or the multistate empirical valence bond (MS-EVB) method used by the Voth group.^{41,42}

EVB couples two or more diabatic states. Where N is the number of diabatic states, the potential energy is given by the lowest eigenvalue of an $N \times N$ Hamiltonian matrix of the type shown in Equation 3. The diagonal matrix elements are the potential energies for each diabatic state calculated by their respective molecular mechanics force fields.

Energy offsets may be added to each diagonal matrix element, calibrated to ensure that the relative energies of the pure diabatic states at their respective equilibrium geometries are correct. The off-diagonal matrix elements are the coupling elements associated with the relevant pairs of states. These can be constants or have a variety of functional forms. In this work they are functions of the reaction coordinate and take the form of Gaussian functions. The Hamiltonian matrix is symmetric such that the two off-diagonal matrix elements associated with a pair of states are identical (e.g. $V_{12} = V_{21}$).

As this is a fairly non-standard method, the reader cannot be pointed to an introductory text explaining the basic principles involved. For this reason, the following description of the development of a prototypical EVB potential has been included in the hope that it is useful to the reader. In this example we build a model for the hypothetical reaction converting A-B + C to A + B-C. In this case all three particles are the same, so a similar scheme could be used to model the reaction of H_A-H_B with H_C. In state 1, A is bonded to B and C has no bonds. The force field describing state 1 has three terms: a bonding term for A-B (as in Equation 1) and two van der Waals' terms describing the non-bonded interactions between A and C, and between B and C (Equation 2). In state 2, B is bonded to C and A has no bonds. Consequently the force field describing state 2 has a bonding term for B-C and two van der Waals' terms describing the interactions between A and B, and between A and C. Our chosen reaction coordinate is the difference of the distances R(A-B) – R(B-C). Our coupling elements V_{12} and V_{21} are identical Gaussians and are functions of the reaction coordinate.

A relaxed scan along the reaction coordinate can be done with some high level *ab initio* quantum mechanics method to generate a reference reaction profile. The EVB and reference surfaces can be compared in several ways. One method is to run single point energy calculations with the EVB potential on the geometries generated with the reference method. The disadvantage of this is that these geometries may not lie along the minimum energy path of the EVB PES. Another method is to perform a relaxed scan with the trial EVB surface and compare the resultant reaction profile to the reference profile. When using this method, care must be taken to ensure that the geometries optimised with EVB are sensible in degrees of freedom other than the reaction coordinate. For example, at $RC = 0$ along the reference minimum energy path, the distances R(A-B) and R(B-C) could be 1 Å, whereas the equivalent distances along the EVB minimum energy path could be 0.5 Å. With this pitfall in mind and having tried both methods, the latter was found to be the more effective method of fitting the EVB potentials described in this

work. Having generated a reference profile and a trial EVB profile, the parameters associated with the EVB potential can be adjusted and a new EVB profile generated. This procedure can be iterated until a good fit is achieved. This can be done either by manually adjusting the parameters and comparing the resultant profiles by eye or by some sort of automated optimisation procedure. In practice, a combination of hand fitting in the early stages and automated fitting for fine-tuning was found to give the best results. The automated procedure involved minimising the sum of the squares of the differences between the EVB and reference energies for all values of the reaction coordinate along the reaction profile. In this example there are a total of 4 force field parameters, 2 offsets and 3 Gaussian parameters, giving a total of 9 parameters, all of which can in principle be adjusted during the fitting procedure. In this case, the reaction is a symmetric reaction, so the parameters ϵ_1 , ϵ_2 and B (the centre of the Gaussian) will all be zero giving a total of 6 adjustable parameters. Experience has shown that beyond this number the fitting procedure becomes increasingly unmanageable using this method.

The resultant EVB profile for our example is shown in Figure 1 alongside the component diabatic states and coupling element.

$$H(q) = \begin{bmatrix} V_{11} + \epsilon_1 & V_{12} \\ V_{21} & V_{22} + \epsilon_2 \end{bmatrix}$$

Equation 3 The Hamiltonian matrix for an EVB PES constructed from two diabatic states. V_{11} and V_{22} are the potential energies calculated with the force fields representing diabatic states 1 and 2 respectively. ϵ_1 and ϵ_2 are the energy offsets for each state. V_{12} and V_{21} are the coupling elements. These are identical and functions of the reaction coordinate.

$$V_{12}(rc) = Ae^{-\frac{(rc-B)^2}{2C^2}}$$

Equation 4 The coupling element V_{12} can have the form of a Gaussian function. V_{12} is a function of the reaction coordinate. The three parameters A , B and C can be adjusted as part of the procedure fitting the EVB surface to reference surface calculated with electronic structure theory.

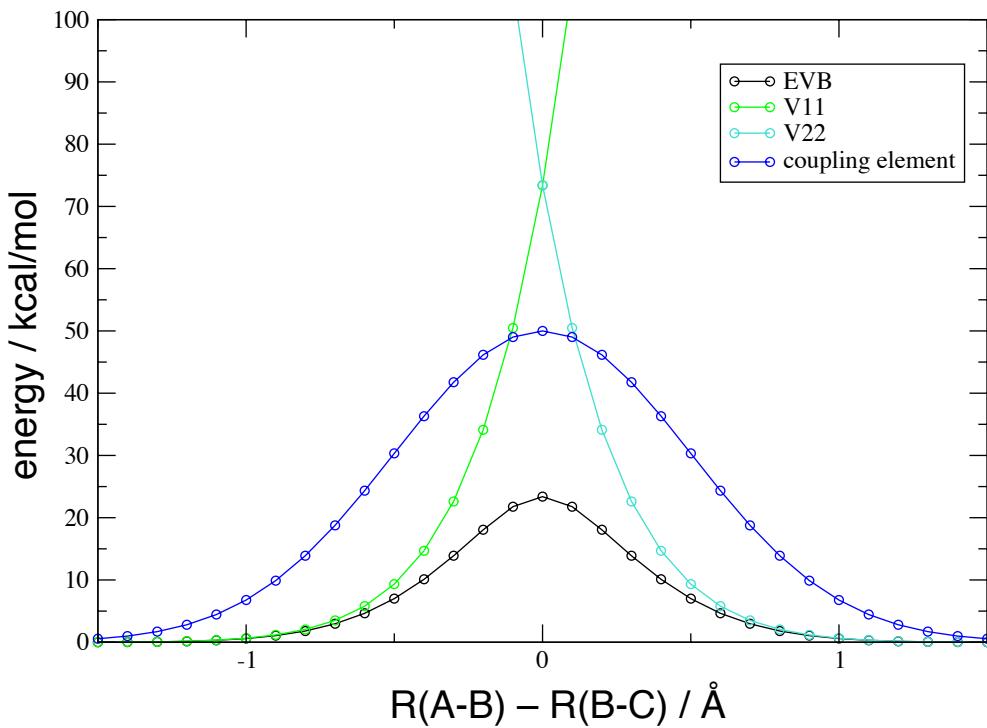


Figure 1 Construction of a prototypical EVB PES (shown in black) from its constituent diabatic states V_{11} and V_{22} . The magnitude of the coupling element (shown in blue) is a function of the reaction coordinate. This surface is for the hypothetical reaction converting $A\text{-}B + C$ to $A + B\text{-}C$.

1.2. Calculating physical properties

1.2.1. Molecular dynamics

Molecular dynamics simulations propagate the positions and velocities of the particles in a system over time by integrating Newton's equations of motion. The resulting trajectories can be analysed to gain insight into physical properties of the system. MD simulations are used by some researchers in order to study time-dependent properties such as diffusion rates or post-reaction product energy deposition.^{40,43} In this work however, MD simulations are used purely as a means to calculate ensemble properties, such as binding free energies, average hydrogen bonding interactions, or the relative populations of particular conformations.

All the simulations of full enzyme systems presented in this work rely on structures from x-ray crystallography as a starting point. While these are immensely valuable in their own right, they present relatively static, time and ensemble averaged structures of enzymes, which often do not properly represent the state of interest. Enzymes are known to fluctuate constantly between vast numbers of interconnected conformations, many of which will be functionally relevant even though they may be infrequently visited.⁴⁴ Given sufficient sampling time, MD simulations can, in principle, thoroughly explore these thermally accessible conformations and provide estimates of their relative populations. At present however, MD simulations of large proteins are generally limited to the sub microsecond regime. Protein conformational changes occur on timescales varying from picoseconds (for dihedral rotation about a low barrier) to milliseconds (for certain folding events). Because of this, it is important always to be aware of the possibility that any given MD simulation may not have sampled the relevant phase space sufficiently well to give reliable insight into a particular problem. As well as adding dynamical information to x-ray structures, MD simulations can model the effects of perturbing a system, for example by mutating an amino acid residue in silico or changing the identity of a ligand bound to an enzyme active site.

1.2.2. Calculating reaction rates

Statistical mechanics provides the tools for prediction of reaction rates from molecular simulation. The transition structure for an elementary step of a reaction is defined as the highest point along the minimum energy path connecting the reactant and product minima on the potential energy surface. Subtly different is the transition state, which is the analogous highest point on the free energy surface.¹ ΔG^\ddagger is the difference in free energy between the transition state and reactant. Classical transition state theory (TST) in its most basic form relates the rate constant for an elementary step of a reaction to the free energy barrier height (ΔG^\ddagger) via the Eyring equation (Equation 5).⁴⁵

$$k = \frac{k_B T}{h} e^{-\Delta G^\ddagger / RT}$$

Equation 5 The Eyring equation relates the rate constant (k) to the free energy difference between the transition state and the reactant (-ΔG[‡]). k_B is the Boltzmann constant, h is the Planck constant, R is the gas constant and T is the temperature.

More rigorous calculations of rates can be performed, for example by multiplying the rate constant obtained by the Eyring equation by a transmission coefficient. This transmission coefficient usually lies in the range 0.5 – 2 and its purpose is to account for the effects of recrossing and tunnelling.⁸ However, as errors in the calculated barrier heights presented in this work will almost certainly affect the rates by factors greater than 2, no transmission coefficients were calculated. Another approximation we have made is to neglect the effect of zero-point energy on reaction rates. Because we are mostly interested in comparing the relative rates of closely related systems, we expect the effects of zero-point energy to cancel out.

It has recently been shown that there exists a quantum transition state theory, capable of calculating the exact quantum rate (in the limit of zero recrossing and assuming the validity of the Born-Oppenheimer approximation) and that this is exactly equivalent to ring polymer molecular dynamics transition state theory (RPMD-TST).^{46,47} Some of the work presented here is likely to be amenable to extension with RPMD-TST simulations, which could, for example, allow for the calculation of kinetic isotope effects on hydrogen transfer reactions in enzymes.

1.2.3. Adiabatic mapping and exponential averaging

There exists a range of sophisticated algorithms for finding accurate transition structures for simple reactions in the gas phase.⁴⁸ Whilst most of these are in principle applicable to QM/MM calculations, they are likely to prove prohibitively expensive for the systems studied in this work (in terms of computational and human resource). We have satisfied ourselves with a more approximate method of finding minimum energy paths and transition structures known as adiabatic mapping. The starting point for an adiabatic mapping calculation is an appropriate reactant structure, usually taken from an MD simulation. What constitutes an appropriate reactant structure is a difficult and important question that will be addressed in the main body of this work. An appropriate reaction coordinate must be chosen that is capable of interconverting the reactant and product structures via a good approximation to the minimum energy path (this choice can be challenging and a potentially significant source of error). A harmonic restraint is placed on the reaction coordinate with a value for the minimum appropriate to the reactant structure. The system is then minimised and the potential energy minus the

restraint energy is recorded along with the value of the reaction coordinate. The minimum for the restraint is then increased (or decreased) towards the product value. This is iterated until a full reaction profile is produced. The potential energy barrier to reaction is then taken to be the difference between the highest point on the surface and the lowest point on the reactant side. Frequency calculations could be performed to confirm the authenticity of the assumed minima and saddle points by checking that they produce exactly 0 and 1 imaginary frequencies respectively. Again we have not done this but bear in mind the errors that this may introduce to our calculated barrier heights.

For simple gas phase reactions at low temperature, the transition structure can be a good approximation of the transition state and reasonably accurate reaction rates can be calculated by substituting the potential energy barrier (ΔE^\ddagger) for ΔG^\ddagger in the Eyring equation.¹ As already discussed, proteins exist in myriad constantly interconverting conformations. Previous experience in our group and others has shown that the potential energy barrier heights obtained from single adiabatic mapping pathways can vary dramatically depending on the starting structure used.⁴⁹ Based on this it is clear that any one of these barriers taken individually is likely to be very poor approximation of the free energy barrier and can consequently lead to gross miscalculations of the rate and unjustified interpretation thereof. In an attempt to better approximate ΔG^\ddagger , multiple adiabatic mapping pathways can be performed starting from a range of structures and the barriers somehow combined. Following from the basic principles of statistical mechanics, the low energy pathways should contribute far more significantly to the barrier heights. Because of this we have chosen here (as in previous work in our group and others) to combine the multiple potential energy barriers by an exponential averaging procedure given in Equation 6.⁵⁰⁻⁵² Though it has been reported in the literature that good results can be obtained by exponential averaging over a relatively small number of pathways, personal experience has shown that convergence of ΔE_{ave}^\ddagger can be slow with respect to the number of pathways.

$$\Delta E_{ave}^\ddagger = -RT \ln \left\{ \frac{1}{n} \sum_{i=1}^n \exp \left(\frac{-\Delta E_i^\ddagger}{RT} \right) \right\}$$

Equation 6 Formula for exponentially averaging the potential energy barrier heights obtained by adiabatic mapping. n is the number of barriers ΔE_i^\ddagger is the barrier height of the i^{th} pathway.⁵²

1.2.4. Calculating free energy barriers

The preference of course would be to directly calculate free energy barriers. The free energy profile along some reaction coordinate can be trivially calculated if the probability distribution along that reaction coordinate is known. For example, the free energy of bringing two molecules in solution into contact can be calculated from the radial distribution function, $g(r)$. The $g(r)$ for these two molecules gives the probability of finding them at a given separation (r) and can be obtain experimentally or calculated from MD (or indeed other types of simulation). The corresponding free energy profile is known as the potential of mean force (PMF) and is calculated via Equation 7.⁵³

$$PMF(r) = -k_B T \ln g(r)$$

Equation 7 Equation relating the potential of mean force (pmf) for the separation of two molecules in solution to the radial distribution function for those particles, $g(r)$.

This method can in principle be used to calculate the PMF along any reaction coordinate. For most reactions of interest however, the barriers are high enough that the regions close to the transition state will essentially never be sampled in an unbiased MD simulation. There are numerous more efficient methods for calculating free energies from simulation.⁵⁴⁻⁵⁸ In this work we have used umbrella sampling combined with the weighted histogram analysis method (WHAM).⁵⁸ Much as with adiabatic mapping, a harmonic restraint is placed along the chosen reaction coordinate at a value appropriate for the reactant. Then, instead of minimising the system, an MD simulation is performed and the value of the reaction coordinate is recorded periodically. The harmonic restraint is increased and further simulation in this new window is performed. This is iterated until the product region is reached. The harmonic restraint enforces sampling of every region of the reaction coordinate in turn. From these simulations, a probability distribution along the reaction coordinate can be built. This will however, be a biased distribution due to the harmonic restraint. The distribution can be unbiased and the free energy profile generated using the WHAM procedure.

Chapter 2. Glutamate mutase

2.1. Introduction

Adenosylcobalamin (AdoCbl), one of the two biologically active forms of vitamin B₁₂, is a cofactor in a small class of enzymes that catalyse reactions via radical mechanisms. The initial steps are common to all enzymes in the class. First is the homolytic cleavage of the AdoCbl Co-C bond to generate a 5'-deoxyadenosyl radical (adenosyl radical) and cob(II)alamin (Figure 2). This is followed (in what may be a concerted or stepwise fashion) by abstraction of a hydrogen by the adenosyl radical from the substrate (Figure 3), or in the case of AdoCbl dependent ribonucleotide reductase, from a polypeptide cysteine.⁵⁹ It has been proposed in the literature that the sole biological purpose of AdoCbl is to act as a source of radicals capable of initiating reactions that may be difficult to catalyse via non-radical mechanisms.⁵⁹ As such, it serves the same purpose as S-adenosyl methionine (SAM) in SAM dependent enzymes (although some of these use the adenosyl radical as a cosubstrate rather than a cofactor).⁶⁰ These enzymes are of interest to study because of their impressive ability to harness the high reactivity of radicals (which are known in solution to react very unspecifically) to catalyse reactions with a very high level of chemo, regio and substrate specificity without damage to the cell often associated with free radicals.

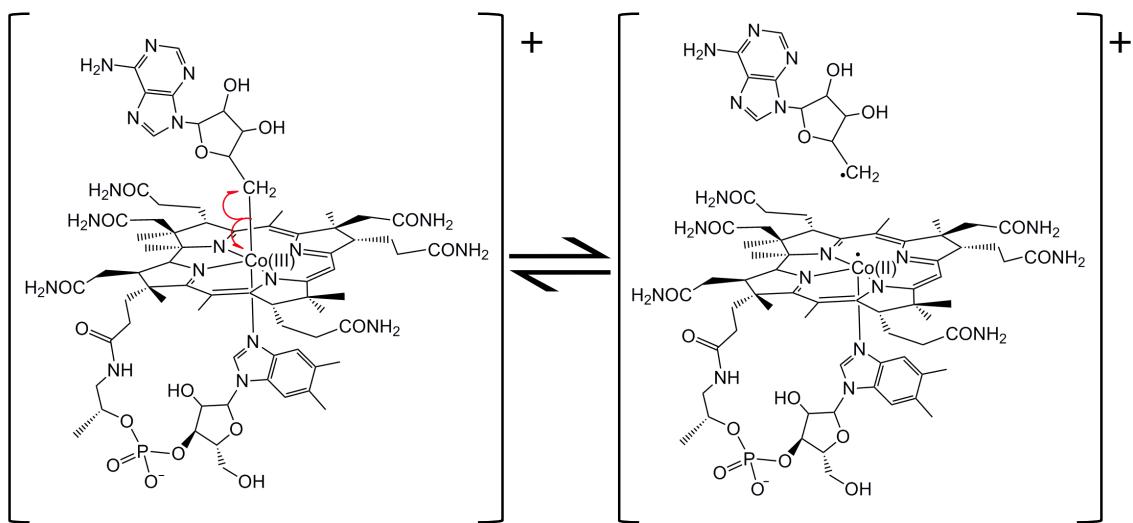


Figure 2 Homolytic cleavage of the Co-C bond in adenosylcobalamin to form the adenosyl radical and cob(II)alamin.

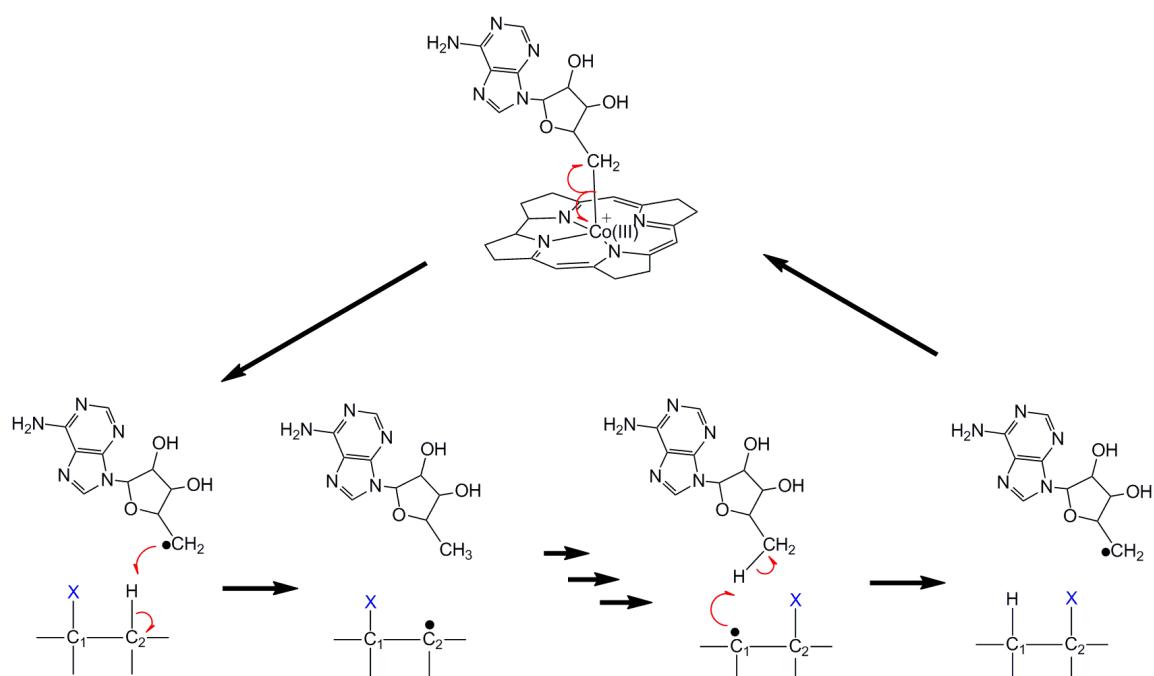


Figure 3 Minimal representation of the reaction mechanism in adenosylcobalamin dependent mutases. The adenosyl radical abstracts a hydrogen from the substrate to form a substrate radical which then rearranges to form a product radical. Re-abstraction of a hydrogen from 5'-deoxyadenosine followed by radical recombination of adenosyl and cob(II)alamin completes the catalytic cycle.

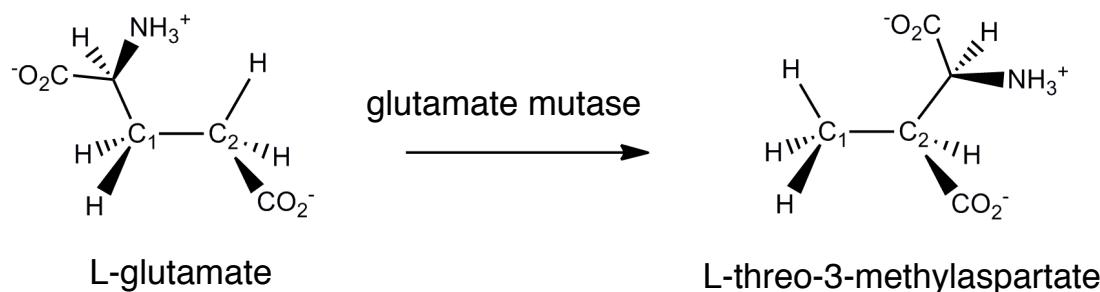


Figure 4 The reaction catalysed by glutamate mutase: the conversion of L-glutamate to L-threo-3-methylaspartate.

AdoCbl dependent enzymes have been (and continue to be) the focus of a significant amount of experimental and computational research. Among the key questions that are being investigated are: how is the Co-C bond dissociation accelerated by an estimated 12 orders of magnitude in enzyme relative to in solution?⁶¹ How are the high-energy radical intermediates controlled to react in such a specific way?⁵⁹ What are the mechanisms of these reactions and what features are common between them?⁶² Does the adenosyl radical exist as a distinct reaction intermediate or as an unstable transient species?^{63,64} How significant is quantum tunnelling of the hydrogen atom in the hydrogen abstraction step and is this a catalytic effect or does tunnelling occur to the same degree in solution as in enzyme?^{65–67}

To explore these questions, we (the Harvey and Mulholland groups in the Centre for Computational Chemistry in Bristol) are engaged in a collaborative project with the experimental group of Professor Neil Marsh at the University of Michigan. For many years Professor Marsh's group has been studying the AdoCbl dependent enzyme Glutamate Mutase (GLM) as a paradigm system for the class as a whole. The natural substrate of GLM is the amino acid L-glutamate, which undergoes a carbon skeleton rearrangement to form L-threo-3-methylaspartate (Figure 4). Using (among other techniques) mutagenesis, isotopic labelling and pre-steady state kinetics and in collaboration with spectroscopists and other computational chemists, the Marsh group has already provided much insight into GLM.^{68–71} Examples include: the elucidation of mechanistic details such as the fragmentation of the glutamyl radical into glycyl radical and acrylate,⁷² the specific roles of certain amino acid residues such as the deprotonation of the substrate amino group by Glutamate 171,⁷³ the kinetic coupling between the Co-C bond dissociation and hydrogen abstraction steps,⁷⁴ the identification of

equilibrium isotope effects and an unusual temperature dependence of kinetic isotope effects indicative of quantum tunnelling in GLM.⁶⁵

With molecular modelling, we aim not only to provide further insight at high temporal and spatial resolution into the existing findings of the Marsh and other experimental groups, but also to make predictions and suggest new ways in which to probe GLM experimentally. In view of this, we have attempted to design our models so that they are capable of:

- modelling the equilibrium properties of the various relevant reaction intermediates,
- calculating reaction profiles and barrier heights for the Co-C bond breaking and hydrogen atom abstraction step(s),
- simulating the reaction in solution (following the same mechanism as in enzyme), so as to identify properties of the enzyme that contribute to catalysis.
- simulating the effects of mutations and other perturbations to the system in experimentally possible ways (such as altering the co-substrate substituents) and non-physical ways (such as altering the charges on active site amino acid residues),
- investigating the nuclear quantum effects of hydrogen tunnelling and zero point energy in order to reproduce (at least qualitatively) experimentally observed kinetic and equilibrium isotope effects

2.2. Review of previous computational studies

DFT, higher level *ab initio* methods and QM/MM have all been used to study the steps involved in rearranging the substrate radical into the product radical (in glutamate mutase and other adenosylcobalamin dependent enzymes).^{71,75,76} The outcomes of these calculations are in good agreement with experiment and these steps are now considered to be well understood.⁶²

In contrast, the details of the Co-C bond breaking and subsequent hydrogen abstraction steps remain open for debate despite the efforts of a large number of experimental and computational groups. A comprehensive overview of this large body of

work will not be given here, instead the reader is directed to an excellent review, discussing progress up to 2009.⁷⁷ For many years, progress in modelling even gas or solution phase Co-C bond breaking was hindered by a lack of consensus about which (if any) DFT functionals are capable of accurately modelling this process. Early studies used B3LYP, which was found to severely underestimate the Co-C bond dissociation energy (BDE) in comparison with experiment.^{61,78} Later, the consensus was that BP86 was the best functional to use because of the good agreement it gave with experimental BDEs.⁷⁹ More recently (after almost all of the studies described here had been conducted) it was shown that there is a large dispersion energy component to the Co-C BDE in these systems which cannot be correctly modelled with normal DFT methods.^{19,80} These studies argued that the most appropriate DFT method to use is B3LYP* (with 15% instead of the 20% Hartree-Fock exchange in regular B3LYP) in conjunction with an empirical dispersion correction (B3LYP*-D).²² Up to the present, the debate concerning how to model cobalamins with DFT continues, with a very recent (August 2014) publication disputing the accuracy of B3LYP*-D in modelling these systems.⁸¹

A DFT (BP86) study of Co-C bond breaking and hydrogen abstraction in a gas phase model found a concerted pathway to be 7 kcal/mol lower in energy than a stepwise mechanism.⁸² Several QM/MM studies have investigated Co-C bond breaking in adenosylcobalamin dependent mutases (including in glutamate mutase).^{63,83–85} These studies favour a stepwise mechanism and propose various factors contribute to facilitating Co-C bond breaking including: electrostatic stabilisation, strain effects, cage effects and even large scale conformational change. Many of these previous studies used adiabatic mapping (see section 1.2.3), which does not allow for significant conformational changes of the active site during the course of the reaction. While this has not hindered the productive application of adiabatic mapping to many systems,^{52,86,87} there are other cases where more thorough sampling of conformational space is required to accurately model reactivity.^{88,89} Because of this drawback and the fact that most of these studies reported only single adiabatic mapping pathways, these studies may suffer severely from a lack of sampling in degrees of freedom orthogonal to the reaction coordinate.

This issue was highlighted in an umbrella sampling EVB study of Co-C bond breaking in methylmalonyl-CoA mutase.⁶⁴ The authors of the study found that significant sampling was required to obtain converged free energy profiles. They argued that none of the adiabatic mapping studies could be accurately describing the Co-C bond breaking process regardless of the accuracy of the potential used. This EVB study found the origin

of the catalytic effect to be the electrostatic stabilisation of the dissociated state via hydrogen bonding interactions of the adenosyl moiety and the protein. Co-C bond breaking and hydrogen transfer were found to be concerted. It is noted that the potentials formed by the particular type of EVB used has come under criticism due to its low degree of emphasis on accurate structure-energy parametrisation, with the focus rather on reproducing experimental properties.⁹⁰

Another study of Methylmalonyl-CoA mutase using umbrella sampling was conducted recently using the Car-Parrinello (BP86) method.³³ In contrast to the EVB study, the Car-Parrinello study found Co-C bond breaking and hydrogen abstraction to occur by a stepwise mechanism. This study also highlighted the importance of sampling, pointing to the fact that significant changes in the conformation of the adenosyl group and its interactions with active site amino acid residues are observed during umbrella sampling (which could not be captured by adiabatic mapping). This study highlighted the importance of the adenosyl -OH groups which (in the dissociate state) hydrogen bond with a conserved active site glutamate. This behaviour has the dual function of electrostatically stabilising the Co-C dissociated state and of guiding the adenosyl radical towards the hydrogen to be abstracted. Having themselves highlighted the importance of sampling in this problem, one drawback of this study is that the amount of sampling performed was fairly limited due to the intrinsic expense of the Car-Parrinello method. 1.5 ps of dynamics was performed in each umbrella sampling window giving a total of 45 ps used to generate an entire free energy profile along the Co-C reaction coordinate. This is a relatively short timescale for exploring some of the degrees of freedom orthogonal to the reaction coordinate, such as dihedral angles. To give an example, one MD study found rotations around dihedral angles in disaccharides to occur on timescales ranging from 30 ps to several nanoseconds.⁹¹ This amount of sampling would certainly not be sufficient to capture any large scale conformational changes that are proposed by some groups to be involved in promoting Co-C bond breaking.^{85,92} This study also suffers from concerns over validity of calculations performed on cobalamin systems using BP86.

Another Car-Parrinello (BLYP) study of methylmalonyl-CoA mutase was published very recently (March 2014).⁹³ This study reveals further conformational changes of the adenosyl group that accompany Co-C bond breaking, again suggesting that this behaviour is involved in both facilitating Co-C bond breaking and carefully directing the adenosyl radical towards hydrogen abstraction.

Chapter 3. Glutamate mutase

PES

3.1. Quantum mechanical PES

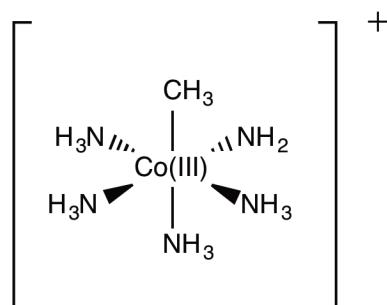
The question of how to model accurately Co-C bond breaking in cobalamins has been a significant on-going challenge for computational chemists for several decades (as discussed in the previous section). The benchmarks for most computational studies have been the experimentally determined solution bond dissociation energies (BDE) for adenosylcobalamin (31.4 ± 1.5 kcal/mol)⁹⁴ and methylcobalamin (37 ± 3 kcal/mol)⁹⁵. The BDEs for these systems calculated with DFT has proved to be highly functional dependent. For example, B3LYP, which was used in many of the earlier DFT studies gives a Co-C BDE for a truncated model of methylcobalamin of 24 kcal/mol (after zero-point energy and solvation corrections), whereas BP86 gives 37 kcal/mol, in apparent excellent agreement with experiment.⁷⁹

One study applying the empirical dispersion correction of the Grimme group to Co-C bond breaking in methyl- and adenosylcobalamin found that dispersion forces contribute 11.3 and 12.8 kcal/mol respectively.²⁰ As BP86 neglects dispersion forces, this result suggests that the good agreement between BP86 and experiment is due to a fortuitous cancellation of errors rather than accurate modelling of the electronic structure. A recent, thorough study was undertaken to determine which functional performs best in calculating the Co-C BDE of the full methylcobalamin cofactor.¹⁹ The calculated BDEs in the study included dispersion, zero-point energy, enthalpy, basis set superposition error, solvation and relativistic corrections. The study concluded that the best available DFT method for studying Co-C bond breaking in cobalamins is B3LYP*-D^a, which gives a Co-C BDE for methylcobalamin of 37.7 kcal/mol, in excellent agreement with experiment. The study also showed that when the Co-C BDE of the full methylcobalamin

^a B3LYP differs from B3LYP in incorporating 15% rather than 20% exact HF exchange.⁹⁶

cofactor is calculated, BP86 is not in as good agreement with experiment as previous studies on truncated systems suggested, giving a Co-C BDE of 30.4 kcal/mol.

In order to reproduce accurately experimental solution phase BDEs, solvent, zero-point energy and enthalpy corrections are required. In an attempt to simplify the problem and to assess more directly the performance of DFT in modelling the electronic structure of Co-C bond containing systems we performed high level *ab initio* BDE calculations on the model system $[\text{CH}_3\text{Co(III)}(\text{NH}_3)_4\text{NH}_2]^+$ (Figure 5).



*Figure 5 Small model of adenosylcobalamin used for high level *ab initio* calculations of Co-C BDE that reasonably mimics the coordination environment of adenosylcobalamin.*

$$\text{BDE} = E[\text{CH}_3] + E[\text{Co(II)}(\text{NH}_3)_4\text{NH}_2]^+ - E[\text{CH}_3\text{Co(III)}(\text{NH}_3)_4\text{NH}_2]^+$$

Equation 8 Equation used to calculate the Co-C BDE. The energy of each fragment $E[\text{fragment}]$ is calculated separately in the gas phase.

The BDE was calculated using Equation 8. The structure of each fragment was optimised using B3LYP and basis set 1 (BS1) using Gaussian 09.⁹⁷⁻⁹⁹ BS1 consists of 6-31G* for H, C and N, and the LanL2DZ basis set for Co, which for Co consists of the Los Alamos effective core potential (ECP) modelling the core electrons and a double zeta basis set for the valence electrons.^{100,101} The ECP includes some relativistic effects through parametrisation. Unrestricted B3LYP (UB3LYP) was used for the open shell calculations on the fragments. The BDE was calculated with various other levels of theory by performing single point energy calculations on the structures optimised with B3LYP/BS1. Dispersion corrections to the DFT energies were added using Grimme's empirical dispersion correction versions 2 (-D2) and 3 (-D3). These were calculated using the standalone dftd3 programme.²² The -D3 correction is more sophisticated and should in principle be used in preference over -D2. However, at the time the calculations were begun, -D3 was unavailable in all of the codes used in rest of this project (QoMMa,

Chemshell, Jaguar, Gaussian and MOLPRO). Because of this we focused on the –D2 correction but have reported the B3LYP*-D3 result for comparison. We found the difference between the –D2 and –D3 corrections to the BDE to be under 2 kcal/mol, so we consider –D2 to model the dispersion forces sufficiently well for our purposes. The dispersion correction parameters were not available for BP86 or B3LYP* in the standalone dftd3 code at the time of calculation so the dispersion correction to the B3LYP energies was added to the BP86 and B3LYP* energies.

CCSD(T)-F12 calculations were performed in MOLPRO^{102,103} using basis set 2 (BS2) which consists of the cc-pVDZ basis set for all C, H and N atoms and the cc-pwCVTZ basis for the Co atom.^{104,105} This basis set is well below the basis set limit, however experience has shown that very good results can be obtained with this size of basis set in combination with the –F12 method.^{14,106} Relativistic effects were added to the coupled cluster calculations with Douglas-Kroll relativistic one-electron integrals (to second order).¹⁰⁷ The ability to calculate Douglas Kroll relativistic corrections with CCSD(T)-F12 has not yet been implemented in MOLPRO. The energy for each fragment was calculated with CCSD(T) with and without relativistic corrections to calculate the CCSD(T) relativistic corrections. These corrections were added to the CCSD(T)-F12 energies.

The results are summarised in Table 1. The performance of the various DFT methods compared to our best estimate of the BDE for this system (calculated by CCSD(T)-F12 with relativistic corrections) is consistent with the performance of these functionals compared to the experimental solution BDE as described in the literature discussed above.¹⁹ BP86 alone (44.59 kcal/mol) does give very good agreement with our best estimate (44.42 kcal/mol), however when the dispersion correction is added, the calculated value is approximately 10 kcal/mol too high. As mentioned before, standard DFT does not model dispersion forces, so the inclusion of the dispersion correction means that the physics of the problem is modelled in a more realistic way. In agreement with Siegbahn²⁰ and Hirao¹⁹, we consider B3LYP*-D2 to be a much better choice for modelling Co-C bond breaking than BP86 as it gives a BDE for this system in as good agreement with our best estimate as BP86 does, but with a more sophisticated description of the physics involved.

Table 1 Co-C bond dissociation energies for the model system shown in Figure 5. Basis set BS1 was used for all DFT calculations and BS2 was used for the coupled cluster calculations.

Level of theory	BDE (kcal/mol)
B3LYP	30.36
B3LYP*	35.45
BP86	44.59
BP86-D2	53.04
B3LYP-D2	38.80
B3LYP*-D2	43.89
B3LYP*-D3	41.98
CCSD(T)-F12	42.36
CCSD(T)-F12 + DK	44.42

3.2. Molecular mechanics PES

In this work we investigated the first two steps of the reaction catalysed by glutamate mutase. We developed three molecular mechanics force fields parametrised to describe the regions of the PES near the three minima associated with these steps (Figure 6). These regions will be referred to as state I - the associated state (with the Co-C bond intact), state II - the dissociated state (with the Co-C bond broken and the radical centre on the adenosyl C5' carbon) and state III - the hydrogen abstracted state (with the Co-C bond broken and the radical centre on the substrate CG carbon). To explore properties of the system in states I and II, molecular dynamics simulations were performed using the relevant unmodified force fields. We refer to these simulations as single state simulations. In order to model the regions of the PES connecting these states, an EVB potential was formed using these force fields and parametrised to fit to QM data as discussed in section 3.4.

The amber99sb force field was used for all standard amino acid residues and ions.²⁶ The 3 point rigid water model TIP3P was used in all simulations.¹⁰⁸ To model adenosylcobalamin we modified published amber compatible force field parameters. The parameters of Marques and Brown were used for the cobalamin group and all bonded

terms between the cobalamin and adenosyl groups^{b,109}. These parameters were carefully fitted to small molecule and enzyme crystal structure data so we considered the bonded parameters to be reasonable. However, we chose not to use the published charges as they differed significantly from amber99sb charges (for example in comparing the amide moieties of the cobalamin to those of glutamine in amber99sb) and because the published charges gave a high concentration of negative partial charges on the atoms surrounding the cobalt atom, which is counterintuitive given its +3 formal oxidation state. The adenine parameters were adapted from published NADH parameters.^{110,111} The free glutamate substrate parameters were adapted from the standard C and N terminal glutamate residues from amber99sb. The charges for the cobalamin, adenosyl, and glutamate substrate were calculated by the RESP fitting method at the HF/6-31G* level of theory using Gaussian09 and RED-IV via the RED server according to the standard amber force field parametrisation protocol.^{99,112} Because of the known poor performance of Hartree–Fock in calculating properties of transition metals, the cobalamin charges close to the Co atom were refined by comparison with Mulliken charges calculated at the B3LYP/LACV3P* level with Jaguar 7.6.¹¹³

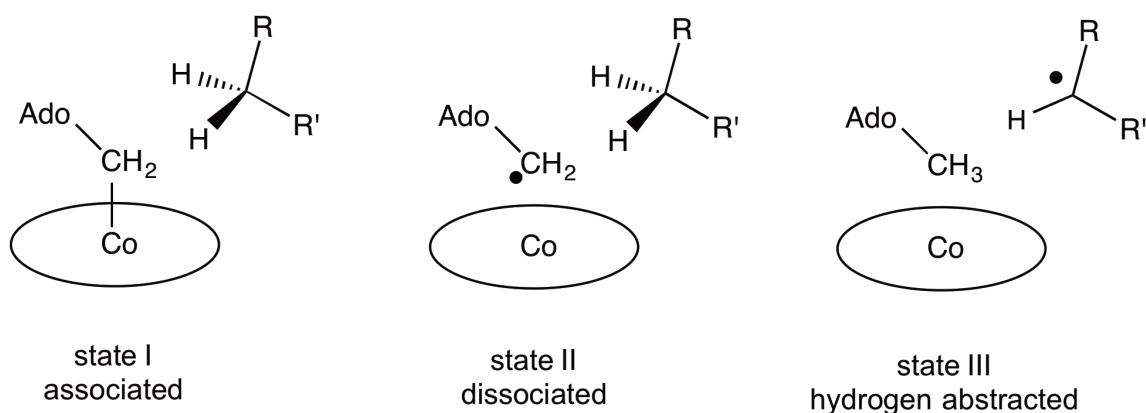


Figure 6 The three states of the glutamate mutase reaction studied in this work. A separate molecular mechanics force field was developed for each state.

State II force field version a (used for the single state MD simulations) differed from state I force field only in the absence of the Co-C bonding and corresponding angle and dihedral terms and the resulting addition of relevant non-bonded terms. As a result of this simplification, some difference in properties between the two states were neglected

^b The supporting information containing the parameters has not been made available online by the publishers so was obtained from Marques in a personal communication.

such as the change in hybridisation of the C5' carbon from sp^3 to sp^2 as it becomes a radical centre and the change in charge distribution on the corrin ring. Repeats of a sample of the single state simulations with version b of this force field showed that the properties of interest (for example the average C5'-HG2 distances as discussed in the next section) obtained from analysis of the single state MD simulations did not significantly change.

Version b of the dissociated state force field (used for the EVB simulations) does model the C5' carbon as a planar sp^2 carbon. This was done by adjusting the angles and dihedrals involving the C5' carbon and including an improper torsion terms H5'-H5''-C4'-C5' and H5''-H5'-C4'-C5'. The parameters for these improper torsion terms were fit to data from a QM relaxed scan of the out of plane bending of the ethyl radical (see Figure 7).

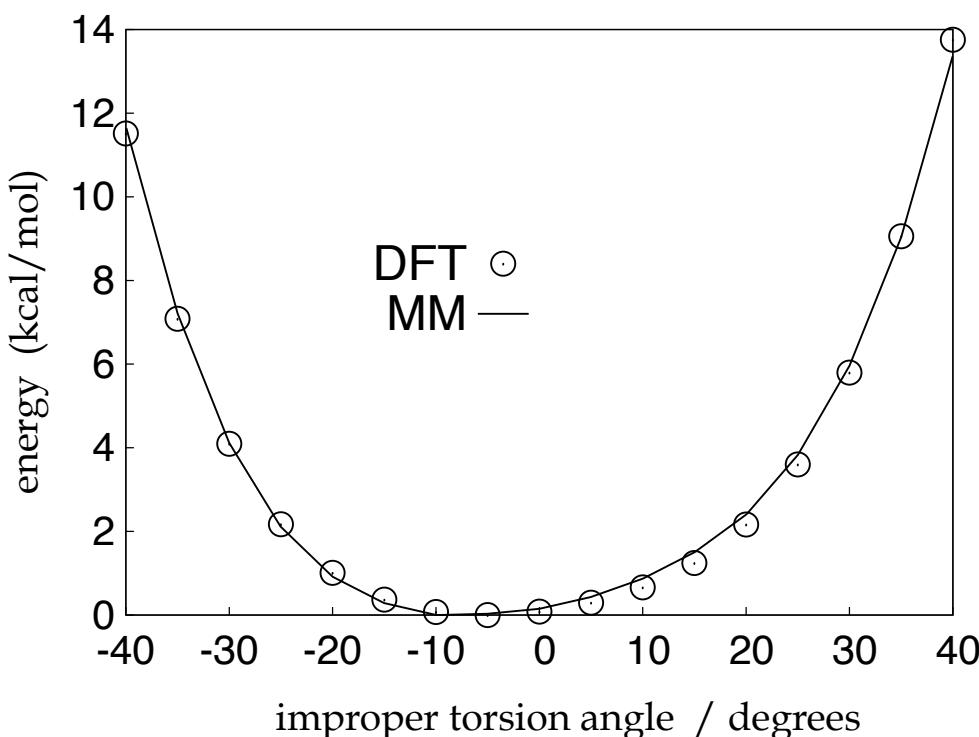


Figure 7 Fitting a molecular mechanics force field to DFT (B3LYP/6-31G*) for the ethyl radical out of plane bending. A DFT relaxed scan was performed of the improper torsion angle H5'-H5''-C4'-C5' (where C5' is the sp^2 carbon). Single point MM energies were calculated on the DFT structures.

In state III, the adenosyl C5' is tetravalent and sp^3 hybridized and the glutamate substrate CG carbon is trivalent and sp^2 hybridized. The sp^2 nature of the CG carbon was

again modelled by adjusting the relevant angles and dihedrals and including the improper torsion terms analogous to those affecting the C5' in the state II force field. The charges on the atoms in the adenosyl fragment and the glutamate substrate change slightly between the dissociated and hydrogen abstracted states to ensure that the total charge on each moiety remains an integer despite the loss or addition of a hydrogen.

3.3. QM/MM PES

QM/MM calculations were run in QoMMa using the protocol outlined in detail in section 1.1.3. B3LYP*-D2/6-31G* was used for the QM calculations. Amber99sb with the Marques cobalamin parameters (as described earlier) was used for the MM calculations. For the calculations of the Co-C bond breaking, the QM region consisted of a truncated cobalamin, imidazole and the adenosyl moiety (Figure 8). For the hydrogen abstraction step, the QM region consisted of the adenosyl radical and the glutamate substrate only. The active region (in which atoms are allowed to move during optimisation) consisted of all whole residues with at least one atom within 6 Å of the QM region. All other atoms were frozen during optimisations.

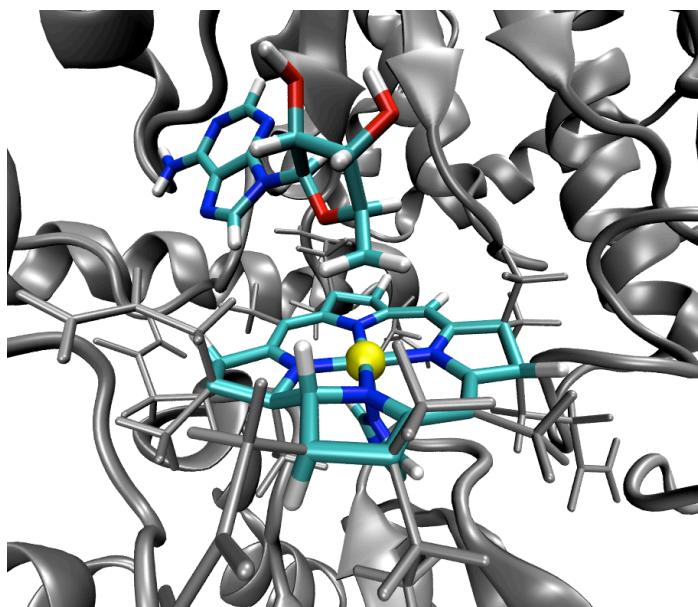


Figure 8 Glutamate mutase active site. Shown in colour is the QM region for QM/MM calculations of the Co-C bond breaking step. The MM region is shown in grey.

3.4. Empirical valence bond (EVB) PES

Separate two state EVB potentials were formed from the three molecular mechanics force fields described in section 3.2 in order to model the Co-C bond breaking and hydrogen abstraction steps. The Hamiltonian matrices for the EVB potentials are shown in Equation 9, Equation 10 and Equation 11. Although most EVB calculations reported here used the two separate two state EVB models, these have been combined to form a three state model (Equation 11) that could be used for further simulations (after further testing). The EVB surfaces were fit to QM data from relaxed scans of the Co-C bond breaking and hydrogen abstraction steps of model systems.

$$H(q) = \begin{bmatrix} V_{11} + \epsilon_1 & V_{12} \\ V_{21} & V_{22} + \epsilon_2 \end{bmatrix}$$

Equation 9 Two state EVB matrix for the Co-C bond breaking step.

$$H(q) = \begin{bmatrix} V_{22} + \epsilon_2 & V_{23} \\ V_{32} & V_{33} + \epsilon_3 \end{bmatrix}$$

Equation 10 Two state EVB matrix for the hydrogen abstraction step.

$$H(q) = \begin{bmatrix} V_{11} + \epsilon_1 & V_{12} & V_{13} \\ V_{21} & V_{22} + \epsilon_2 & V_{23} \\ V_{31} & V_{32} & V_{33} + \epsilon_3 \end{bmatrix}$$

Equation 11 Three state EVB matrix for the Co-C bond breaking and hydrogen abstraction steps.

A truncated model of adenosylcobalamin (shown in Figure 9) was used for the fitting of the EVB potential for the Co-C bond breaking step. A relaxed scan of the Co-C bond was performed at the B3LYP*/6-31G* level of theory using Jaguar.¹¹³ The Co-C bond was held fixed while all other atoms were fully optimized. The value of the Co-C bond length was increased from 1.8 Å to 4.0 Å in 0.1 Å increments. Single point dispersion corrections were added. These were calculated with Grimme's stand alone DFTD3 programme using version 2 of the dispersion correction.^{22,114} As dispersion

correction parameters were not available in the code used for BP86 and B3LYP*, the dispersion corrections calculated for B3LYP were used to correct all functionals discussed here.

Relaxed scans were performed in CHARMM using a two state EVB model comprised of the state I and II force fields. The objective of the fitting procedure was to achieve a good fit between the reaction profiles obtained by QM and EVB relaxed scans by adjusting the parameters associated with the coupling elements, offsets and force fields. Prior experience has shown that fitting an EVB potential to reference data is greatly facilitated if the component force field potentials (the diagonal elements of the EVB matrix) match the reference data well near their respective minima and if the crossing point for these potentials is not too high in energy. In light of this, some force field parameters were adjusted to meet these criteria before attempting to adjust the coupling element parameters. These were the force constant for the Co-C bond stretching term in the state I force field and the Lennard-Jones parameters for the Co-C non-bonded interaction in the state II force field.^c Version b of both state I and state II force fields were used in the final EVB potential.

The sum of two Gaussian functions (which are functions of the Co-C bond distance) form the coupling elements V_{12} and V_{21} (note that $V_{12} = V_{21}$) as shown in Equation 12. The EVB and QM relaxed scans are shown in Figure 10.

$$V_{ij}(rc) = A_{ija} e^{-\frac{(rc - B_{ija})^2}{2C_{ija}^2}} + A_{ijb} e^{-\frac{(rc - B_{ijb})^2}{2C_{ijb}^2}}$$

Equation 12 $V_{ij}(rc)$ is the coupling element between states i and j as defined in Equation 9-4. These coupling elements are linear combinations of Gaussian functions dependant on the reaction coordinate (rc). The parameters to be fitted are the amplitude (A), centre (B) and width (C) of each Gaussian. rc is the Co-C distance for V_{12} and $R(C5'-H) - R(CG-H)$ for V_{23} .

^c As standard, CHARMM derives the Lennard-Jones parameters for a given pair of atoms from the homonuclear diatomic Lennard-Jones parameters of the individual atom types. However, it is also possible to specify the Lennard-Jones parameters for any particular pair of atom types using the NBFIX keyword in the parameter file, thereby allowing for the adjustment of the Lennard-Jones interaction between a particular pair of atom types without affecting any other interactions. The Lennard-Jones interactions were adjusted here using NBFIX.

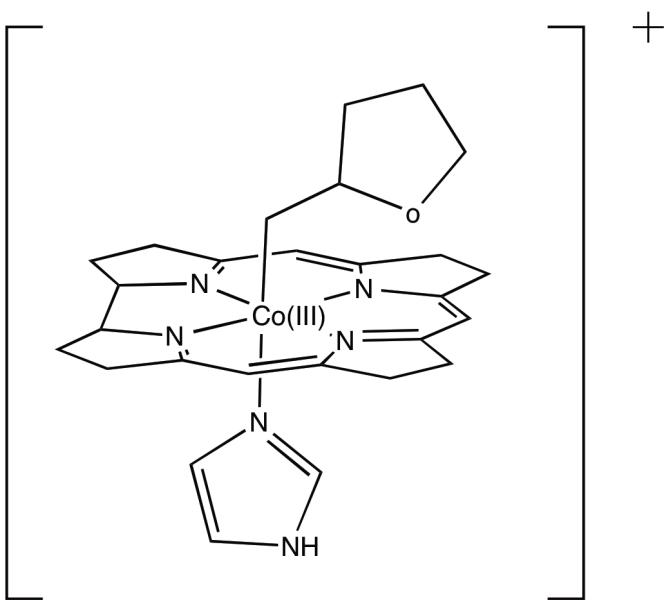


Figure 9 Model of adenosylcobalamin used for fitting the EVB potential for the Co-C bond breaking step

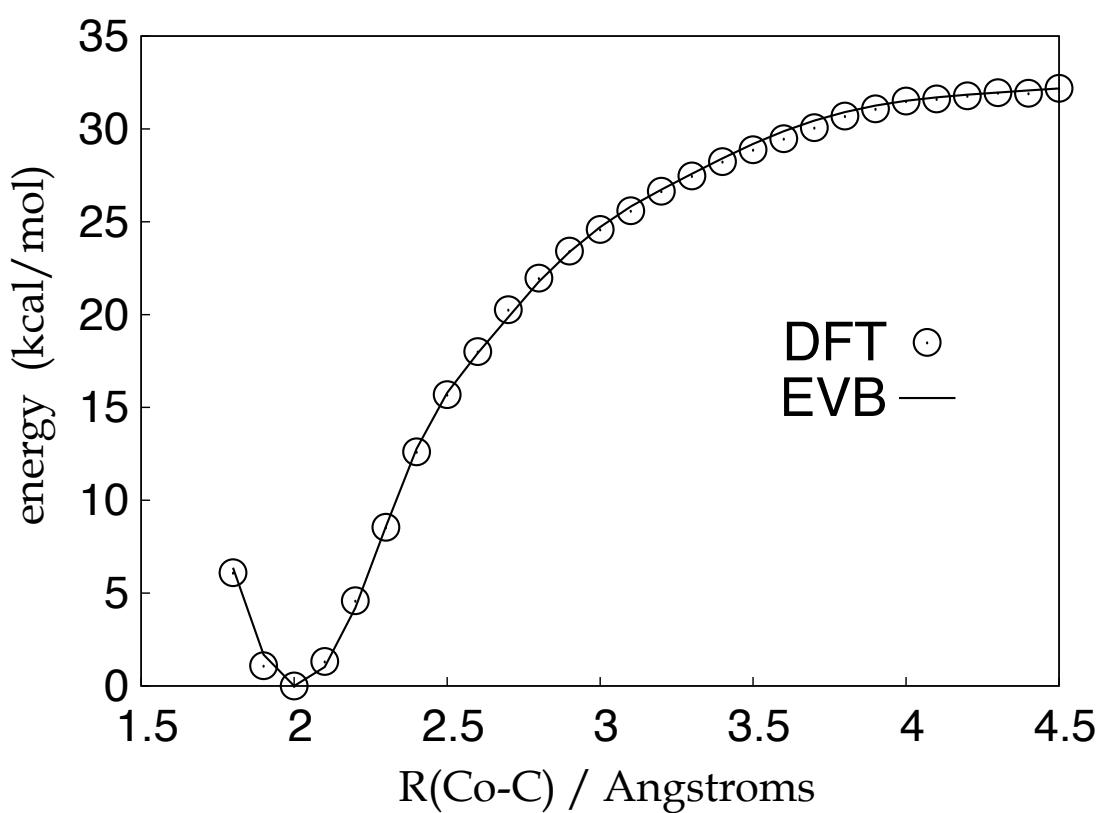


Figure 10 Relaxed scans of Co-C bond breaking in model system performed with EVB and DFT.

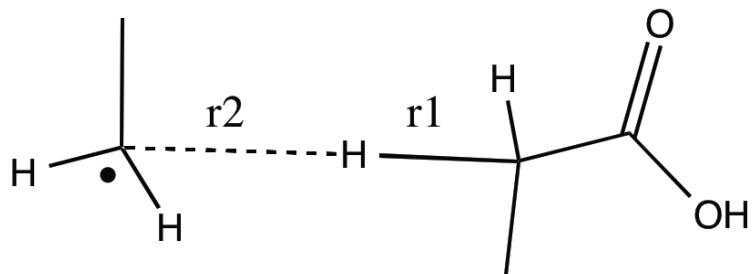


Figure 11 Model used for EVB fitting of hydrogen abstraction step. The reaction coordinate was $R(r_1 - r_2)$.

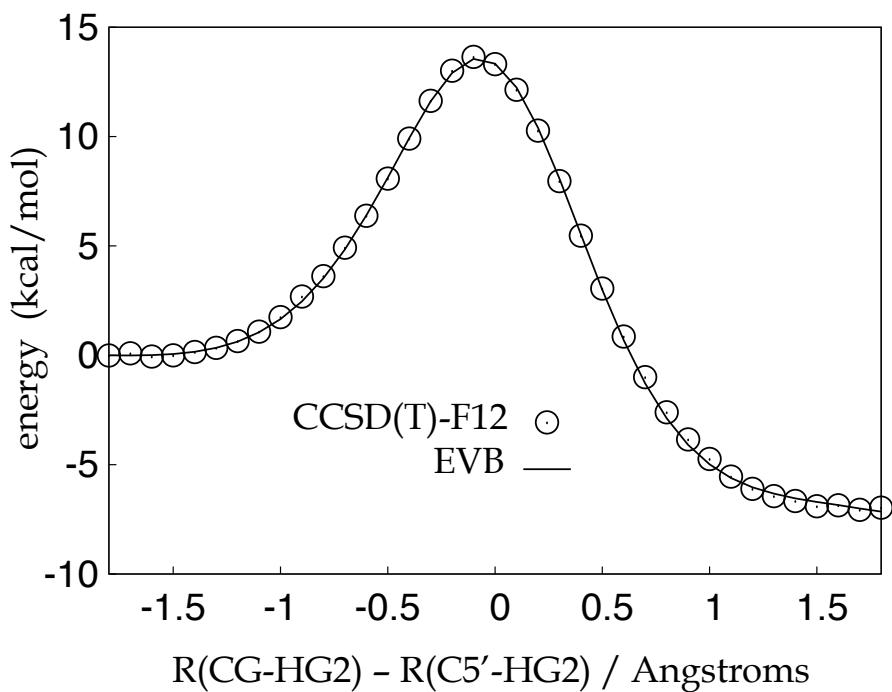


Figure 12 Relaxed scans of H abstraction step in model system shown in Figure 11.

The model of adenosyl and glutamate shown in Figure 11 was used for the fitting of the EVB potential for the hydrogen abstraction step. A relaxed scan of the difference of distances reaction coordinate $R(C5'-H) - R(CG-H)$ at 0.1 Å increments was performed at the B3LYP-D2/6-31G level of theory using ChemShell interfaced with Gaussian 09.^{99,115} Single point energies were calculated at each point at the CCSD(T)-F12/aug-cc-pvdz level of theory using MOLPRO. Once again the main objective of the fitting procedure for this step was to achieve a good fit between the reaction profiles obtained by QM and EVB relaxed scans. In addition to this we aimed to get a good match for the bond lengths $R(C5'-H)$ and $R(CG-H)$ between the EVB potential and reference QM potential at the transition state.

To model the hydrogen abstraction step, a two state EVB potential comprised of the force fields for states II and III was formed. By simply adjusting the Lennard-Jones parameters we were unable to get the component force fields (states II and III) to model simultaneously the potential well at their respective minima, cross at acceptably low energy and give correct bond lengths at the transition state and minima. We identified the steepness of the repulsive component of the Lennard-Jones potential as being the source of this difficulty. To reduce the steepness of this repulsion we used the separation-shifted scaling soft core van der Waals and electrostatic potentials of McCammon for the non-bonded interactions between the three atoms involved in the reaction coordinate only (no other interactions were affected).¹¹⁶ This method is implemented in CHARMM via the BLOCK and PSSP modules.

$$V_{elec} = \frac{q_i q_j}{r_{ij}}$$

Equation 13 Standard amber and CHARMM force field electrostatic term.

$$V_{elec} = \lambda \left[\frac{q_i q_j}{\sqrt{r^2 + \delta(1-\lambda)}} \right] + (1-\lambda) \left[\frac{q_i q_j}{\sqrt{r^2 + \delta\lambda}} \right]$$

Equation 14 Separation shifted soft core electrostatic potential. The parameters used in this study are $\lambda = 0.99$, $\delta = 350$.

$$V_{LJ}(r) = \frac{A}{r^{12}} - \frac{B}{r^6},$$

$$A = 4\epsilon\sigma^{12}, \quad B = 4\epsilon\sigma^6$$

Equation 15 Standard Lennard-Jones potential (AB form).

$$V_{LJ} = \lambda \left[\frac{A}{(r^2 + \delta(1 - \lambda))^6} - \frac{B}{(r^2 + \delta(1 - \lambda))^3} \right] + (1 - \lambda) \left[\frac{A}{(r^2 + \delta\lambda)^6} - \frac{B}{(r^2 + \delta\lambda)^3} \right],$$

$$A = 4\epsilon\sigma^{12}, \quad B = 4\epsilon\sigma^6$$

Equation 16 Separation shifted soft core vdW's potential. The parameters used in this study are $\lambda = 0.99$, $\delta = 350$.

Table 2 Parameters for the EVB coupling elements as defined in Equation 12. These coupling elements are used in Equation 9, Equation 10 and Equation 11. The A parameters have units of kcal/mol. The B and C parameters have units of Angstroms.

rc		A _{ija}	B _{ija}	C _{ija}	A _{ijb}	B _{ijb}	C _{ijb}
V ₁₂	R(Co-C)	9	2.601	0.274	36	3.243	0.579
V ₂₃	R(CG-H) - R(C5'-H)	56	0	1000	13	-0.08	0.42
V ₁₃	-	0	-	-	0	-	-

Table 3 EVB offset parameters as used in Equation 9, Equation 10 and Equation 11. All offsets are in kcal/mol.

ϵ_1	ϵ_2	ϵ_3
0	20.3	7.4

Chapter 4. Characterising the Co-C associated and dissociated states

4.1. Introduction

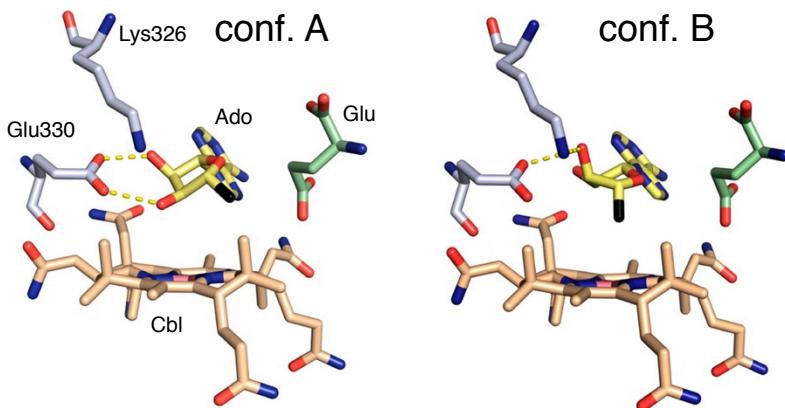


Figure 13 Two distinct conformations of the adenosyl moiety are given in the crystal structure. These labels A and B are taken from the pdb file. Note that the labels are reversed in the original crystal structure paper.¹¹⁷ Glu330 and Lys326 were the residues targeted for mutation. The glutamate substrate is labelled Glu. Figure adapted from a paper published by the author and collaborators.¹¹⁸

How Co-C bond breaking is facilitated and the resulting high energy adenosyl radical species controlled are two of the most fundamental questions about adenosylcobalamin chemistry and both remain unsatisfactorily answered. One theme that is beginning to emerge from numerous experimental and computational studies on various different adenosylcobalamin dependent enzymes is that significant conformational changes are likely to occur at various points during turnover. Proposed functionally relevant phenomena include: changes in adenosyl to enzyme hydrogen bonding¹¹⁹, pseudorotation of the adenosyl ribose group¹¹⁷, substrate binding induced changes to active site structure¹²⁰, changes in adenosyl internal dihedral conformations⁹³ and even large scale protein domain motions.⁸⁵

The main aim of the work described in this chapter is to characterise two points along the reaction cycle (the Co-C associated and dissociated states) and to identify potentially functionally relevant structural differences between them. In running simulations of the dissociated state using a force field with no Co-C bonding term we are of course making the assumption that the adenosyl radical does exist as a distinct intermediate species. The validity of this assumption will be tested with EVB simulations of the reaction in the following chapter.

As discussed in 1.2.1, structures from x-ray crystallography present a time and ensemble averaged picture of enzyme structure and may not accurately represent a functionally relevant state. The crystal structure of glutamate mutase (pdb code 1I9C) clearly does not represent the system in the associated state (with Co-C bond formed). A superposition of two adenosyl conformations was found to best fit the electron density (Figure 13). Conformations A and B had an occupation ratio of 60:40. In neither of these conformations was the Co-C bond intact, having Co-C distances of 4.2 Å and 3.2 Å respectively (in contrast, crystal structures of adenosylcobalamin out of enzyme have Co-C bond lengths of around 2 Å¹²¹). It is possible that one or both of these conformations do accurately represent the dissociated state, in which case fairly short MD simulations should provide reliable estimates of structural properties. In contrast, if the conformational changes between the associated and dissociated states are as significant as is hinted in some of the literature, much longer MD simulations will be required to characterise the associated state properly. One major criticism of previous computational studies on Co-C bond breaking in enzyme is that few if any of the other studies had run MD simulations long enough to allow for proper equilibration of the associated state (on the basis of the conclusions of this chapter).^{33,84,122}

Running in parallel to this computational work, our experimental collaborators in the Marsh group at the University of Michigan were conducting kinetic experiments on glutamate mutase. Based on previous computational work^{76,123}, x-ray structures^{117,120} and preliminary simulations from this work, two residues were identified as potentially important in stabilising the dissociated state and in controlling the adenosyl radical via hydrogen bonding to the adenosyl –OH groups. These were Glu330 and Lys326 (see Figure 13). The mutants E330A, E330D, E330Q, K326M and K326Q were all expressed and their activity probed by monitoring the rate of tritium exchange between radiolabelled glutamate and adenosylcobalamin. A range of diminished activity relative to the wild type was observed for these mutants. Because of the kinetic coupling between

the Co-C bond breaking steps and hydrogen abstraction, it is unclear from the experiments alone whether the observed reduced activity is due to a reduction in the rate of Co-C bond breaking or hydrogen abstraction or both. In an attempt to rationalise the experimental results, we have studied all of these mutants computationally alongside the wild type. Some of this joint experimental and computational work has been published and parts of the manuscript adapted for this chapter.¹¹⁸

4.2. Computational details

4.2.1. System preparation

The starting point for all simulations performed in this study was the structure of glutamate mutase from clostridium cochlearium in complex with adenosylcobalamin and substrate obtained by x-ray crystallography (PDB code 1I9C).¹¹⁷ The entire biologically active tetramer (containing two active sites) was simulated. Both conformations A and B of the adenosyl group were used as starting points for MD simulations (Figure 13). Where alternate conformations exist for other residues in the crystal structure, the conformation with the higher occupation number was used.

The resolution for the crystal structure is 1.90 Å, which (as with most protein crystal structures) is not high enough to resolve the positions of hydrogen atoms. Consequently all hydrogens had to be added in silico. The pKas of ionisable residues were calculated with the web server PROPKA.¹²⁴ Protonation states of histidine residues were determined based on these calculated pKas and by visual inspection of implied hydrogen bonds. The calculated pKas showed all other ionisable residues to be in their standard protonation states at pH 7. Some Asn and Gln residues were flipped (the positions of OD1 and ND2 exchanged) to remove steric clashes as calculated with the structure validation programme Molprobity.¹²⁵

The protein was solvated in an orthorhombic water box with > 12 Å between the protein and the sides; sodium ions were added to neutralize the system.¹²⁶

Structures for mutant systems were generated using the ‘mutagenesis wizard’ tool in the PyMol molecular graphics system, version 1.3 (Schrödinger, LLC).

For simulations of the associated state of adenosylcobalamin in solution, the entire axial histidine residue was included to complete the coordination of the cobalt. Note that

this does not exactly correspond to the state of adenosylcobalamin in solution in which it is known that the axial ligand is either the 5,6-dimethylbenzimidazole base (part of adenosylcobalamin itself) or a solvent molecule.⁹⁴ For the solution simulations of the dissociated state only the adenosyl radical was modelled. The solution systems were solvated in an octahedral water box with > 12 Å between the solute and the sides.

The AMBER ff99SB force field was used for all standard protein residues, ions and water.²⁶ The parameters used for the adenosylcobalamin, adenosyl radical and glutamate substrate are described in section 3.2.

The AmberTools 12 program ‘tleap’ was used to add the hydrogen atoms, solvent and ions and to build the topology and parameter files for use with the Amber MD simulation package.¹²⁶ The total number of atoms in each simulation was approximately 140000.

4.2.2. Molecular dynamics simulations

The simulation procedure described below was used for all systems except where mentioned otherwise. The positions of all water molecules and ions were optimized with 100 steps of steepest descent followed by 900 steps of conjugate gradient minimization. The whole system was then similarly optimized with restraints of 5 kcal mol⁻¹ Å² on all α-carbons. Energy minimizations were performed using the sander.MPI code in the AMBER11 package. Random velocities were assigned, and the system was heated to 300 K by 50 ps Langevin dynamics (collision frequency of 5 ps⁻¹) with α-carbon restraints applied; a 100 ps equilibration in the *NPT* ensemble at 300 K and 1 atm was performed with α-carbon restraints present followed by the gradual release of restraints after additional simulation over 50 ps. Production runs were performed using Langevin dynamics (collision frequency of 5 ps⁻¹) at 300 K and 1 atm. All MD simulations were performed with a 2 fs time step. The SHAKE algorithm was applied to constrain the length of all bonds involving hydrogen, allowing a time step on a time scale similar to that of vibrations of bonds involving hydrogens to be used. Nonbonded interactions were not calculated between pairs of atoms separated by a distance greater than a cutoff of 10 Å. Periodic boundary conditions were used with an orthorhombic box for the enzyme systems and an truncated octahedron for the solution systems. The particle mesh Ewald

(PME) method was applied to treat long-range electrostatic interactions. All MD simulations were performed using the pmemd.mpi code in the AMBER11 package.¹²⁶

Only one simulation was performed for each enzyme system. However, as each contains two active sites, for the purposes of most of the analysis reported here this is largely equivalent to having run two simulations for each system. All of the analysis of active site properties was performed on the combined data from both active sites for each system. Two separate 20ns runs were performed for each solution system.

All analysis was performed on trajectories comprised of structures saved every 20ps from the production phases. Unless otherwise stated, the MD trajectories were analysed using the AmberTools 12 programme ‘cpptraj’ (version 12.5).¹²⁶

Table 4 Summary of MD simulations. Each enzyme system was simulated in the associated and dissociated states for the length of time indicated.

system	adenosyl starting conformation	length (ns)
wild type	A	20
wild type	B	20
wild type (no substrate)	A	20
solution (AdoCbl)	A	40
solution (Ado only)	A	40
E330A	A	15
E330D	A	15
E330Q	A	15
K326M	A	15
K326Q	A	15

4.3. Results and discussion

4.3.1. Active site structure

Figure 14 shows the superposition of average structures from MD simulations of the associated and dissociated states on top of the crystal structure (with adenosyl conformation A). The average structure of the dissociated state simulation is extremely similar to the crystal structure. We interpret this as evidence that the crystal structure with adenosyl in conformation A is representative of the dissociated state. Furthermore it

suggests that the force field used models the structural properties of this state well. In contrast, the average structure from the associated state simulation differs significantly from the crystal structure, particularly in the internal conformation of the adenosyl group and its hydrogen bonding interactions with Glu330, Lys326, Asn123 and Gly68.

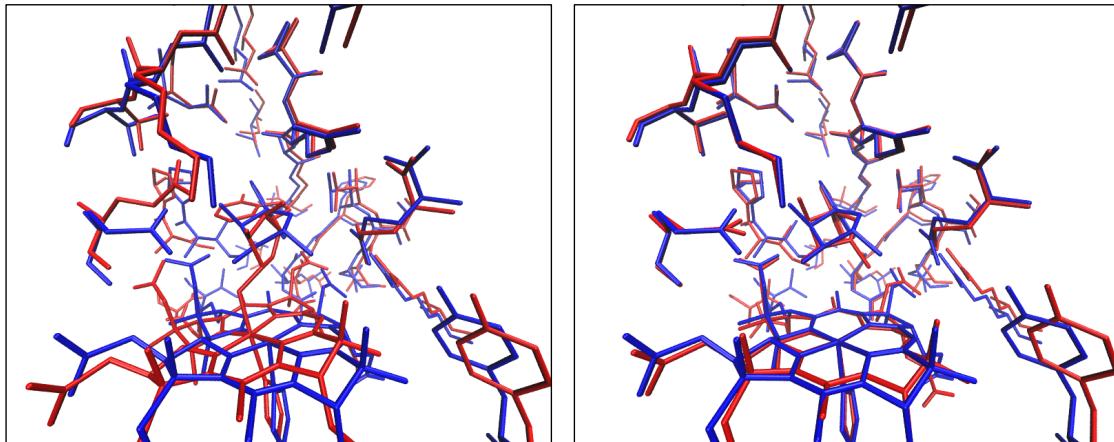


Figure 14 Superposition of the average structures from single MD simulations (red) of the associated (left) and dissociated (right) states on top of the crystal structure (with the adenosyl group in conformation A). Structures were aligned so as to minimize the RMSD between the heavy atoms of residues within 6 Å of the adenosyl moiety.

The root-mean-square-deviation (RMSD) of the positions of the C α backbone carbon atoms relative to the crystal structure was calculated for each frame of the production trajectories. This was done with VMD version 1.9.1 using custom written scripts.¹²⁷ The RMSDs were calculated for C α atoms from every residue as well as for C α atoms belonging to active site residues only. Active site residues were defined as any residue containing at least one atom within 6 Å of the adenosyl moiety in the crystal structure (a total of 24 residues per active site). All frames from each trajectory were aligned with the crystal structure so as to minimize the overall C α RMSDs. This alignment procedure was repeated to minimize the RMSDs of the active site C α . Figure 15 shows the RMSD results for the wild type simulations. The RMSDs for the dissociated state are low and remain constant throughout the simulations. The RMSDs for the associated state simulations are higher than for the dissociated state. This indicates that the enzyme backbone structure is significantly different in the associated and dissociated states. The RMSDs for the associated state are also less stable and appear to still be increasing towards the end of the simulation, indicating that we may not have

arrived at an accurate representation of the associated state even by the end of the simulations.

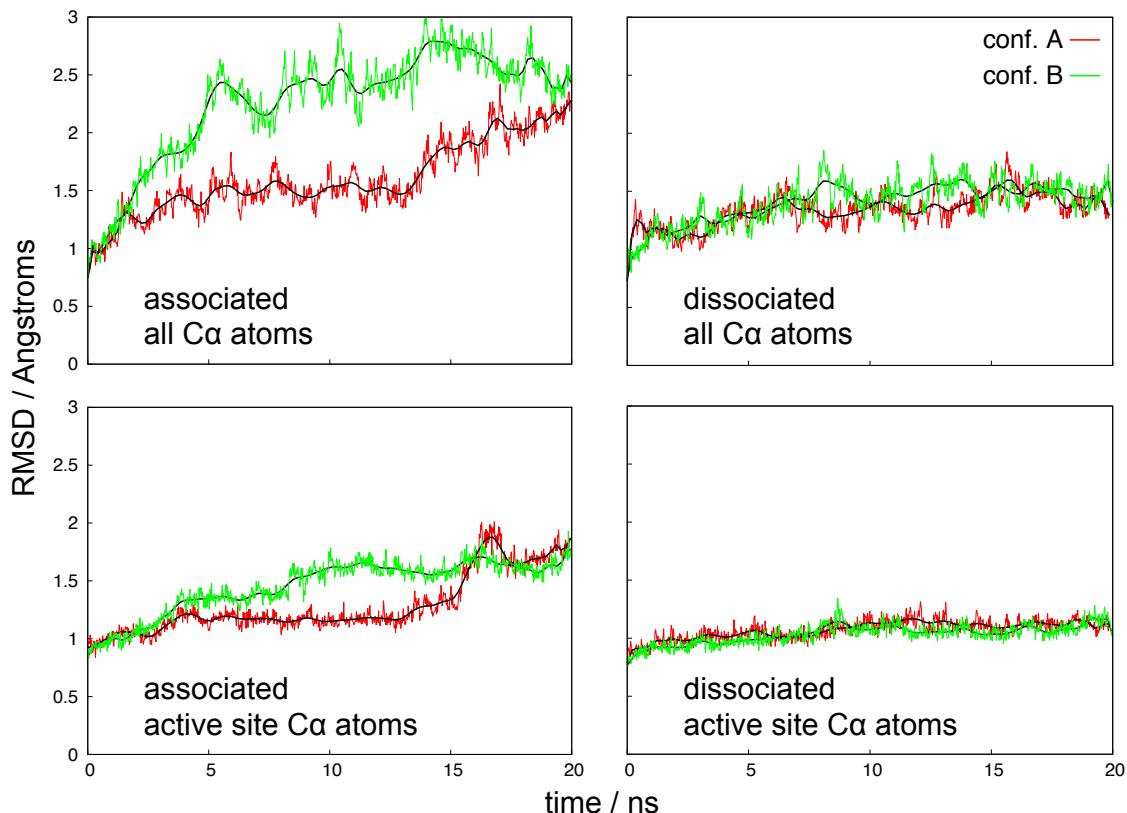


Figure 15 RMSDs over time of Ca backbone atoms with reference to the crystal structure. Active site residues are defined as any residue containing at least one atom within 6 Å of the adenosyl moiety in the crystal structure. The lines in black are Bézier curves fit to the raw RMSD data using gnuplot 4.6.

Figure 16 shows the results for wild type, wild type without substrate and mutant enzymes. The active site RMSDs (but not the global RMSDs) for the wild type without substrate are significantly higher than for the wild type. This is consistent with their hypothesis that conformational change accompanies substrate binding. On its own however, this does not give any indication of the significance of this conformational change. Indeed it would be surprising if little active site rearrangement was observed given that the substrate clearly forms numerous salt bridge and hydrogen bonding interactions with active site residues (substrate binding is not discussed extensively here, for more details see the following references^{76,117}). The RMSDs for some of the mutant

systems in both states are considerably higher and less stable than those for the wild type. This indicates that the introduction of these mutations does significantly alter the structure of the active site and beyond. When interpreting further results for these systems it should be borne in mind that the enzyme structures may not have had sufficient opportunity to adjust to the perturbation introduced by the mutations.

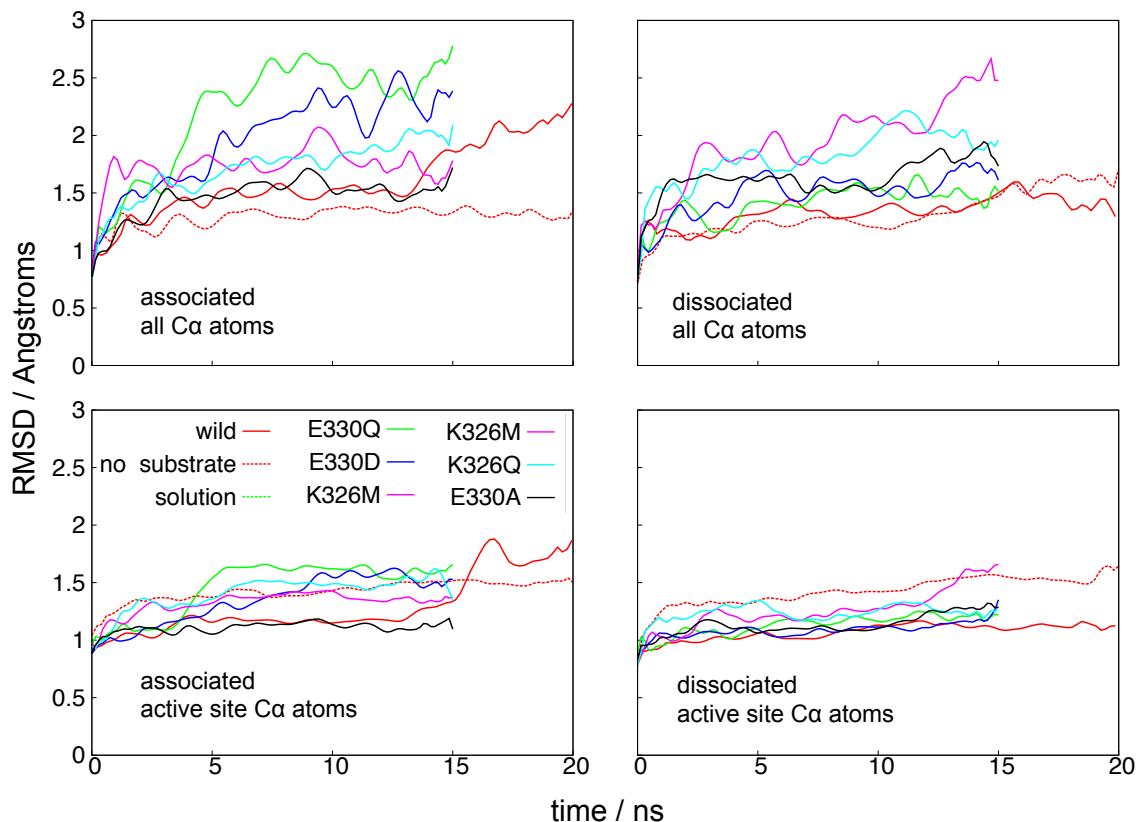


Figure 16 RMSDs over time of Ca backbone atoms with reference to the crystal structure for the wild type, wild type with no substrate and mutant enzymes. Active site residues are defined as any residue containing at least one atom within 6 Å of the adenosyl moiety in the crystal structure. All simulations shown here were started with the adenosyl moiety in conformation A. The lines are Bézier curves fit to the raw RMSD data using gnuplot 4.6.

4.3.2. Conformations of the adenosyl moiety

It is evident from visualisation of the wild type trajectories that the associated and dissociated states differ significantly with respect to the conformation of the adenosyl moiety and its hydrogen bonding interactions with the protein. Five structural parameters were used to analyse the adenosyl conformations: the puckering amplitude and pseudorotational phase angle of the ribose group and the dihedral angles O4'-C1'-N9-C4A (Φ), Co-C4'-C5'-O4' (θ) and N24-Co-C5'-C4' (χ) illustrated in Figure

17. The average values and standard deviations of these dihedral angles from the wild type simulations and their values in the crystal structure are given in Table 5. The distributions of these dihedral angles are shown in Figure 18.

The average values for the dihedral angles from the full length simulations of the dissociate state are similar to those in conformation A of the crystal structure. By looking at the standard deviations and the histograms of the distributions, it is clear that the conformation of the adenosyl group in the dissociated state remains stable and close to crystal structure conformation A throughout.

In contrast, the average values from the full length simulations of the associate state are very different from either crystal structure conformation and the corresponding standard deviations are high. All three of the dihedral angles for the full length simulations of the associated state clearly have bimodal distributions. If only the last 5 ns of each trajectory is analysed however, a single peak emerges in each case. The traces of these dihedral angles over time for the wild type in the associated state are shown in Figure 19. These traces show that the adenosyl moiety tends towards the same internal conformation in all 4 cases (in both active sites starting from both adenosyl crystal structure conformations). Looking at the traces for the simulations started in conformation A it appears that the adenosyl moiety is trapped in some conformation in the beginning of the simulation, likely as an artefact of the starting structure. At some point in both active sites, the adenosyl moiety relaxes into another conformation. In the case of the associated state simulations starting from adenosyl conformation B, values of χ and θ have arrived near their end values at some point during equilibration (not shown on the traces). The dihedral Φ however, behaves as it did for the simulations starting in conformation A. Once close to the final values, they close for the rest of the simulation. Based on this, we assume that the conformations arrived at by the end of each simulation are lower in energy than earlier conformations. The average values of the dihedral angles from the last 5 ns of simulation give our best estimate of the adenosyl conformation in the equilibrated associated state, although we cannot rule out the possibility that if longer simulations were performed, further conformational changes might occur.

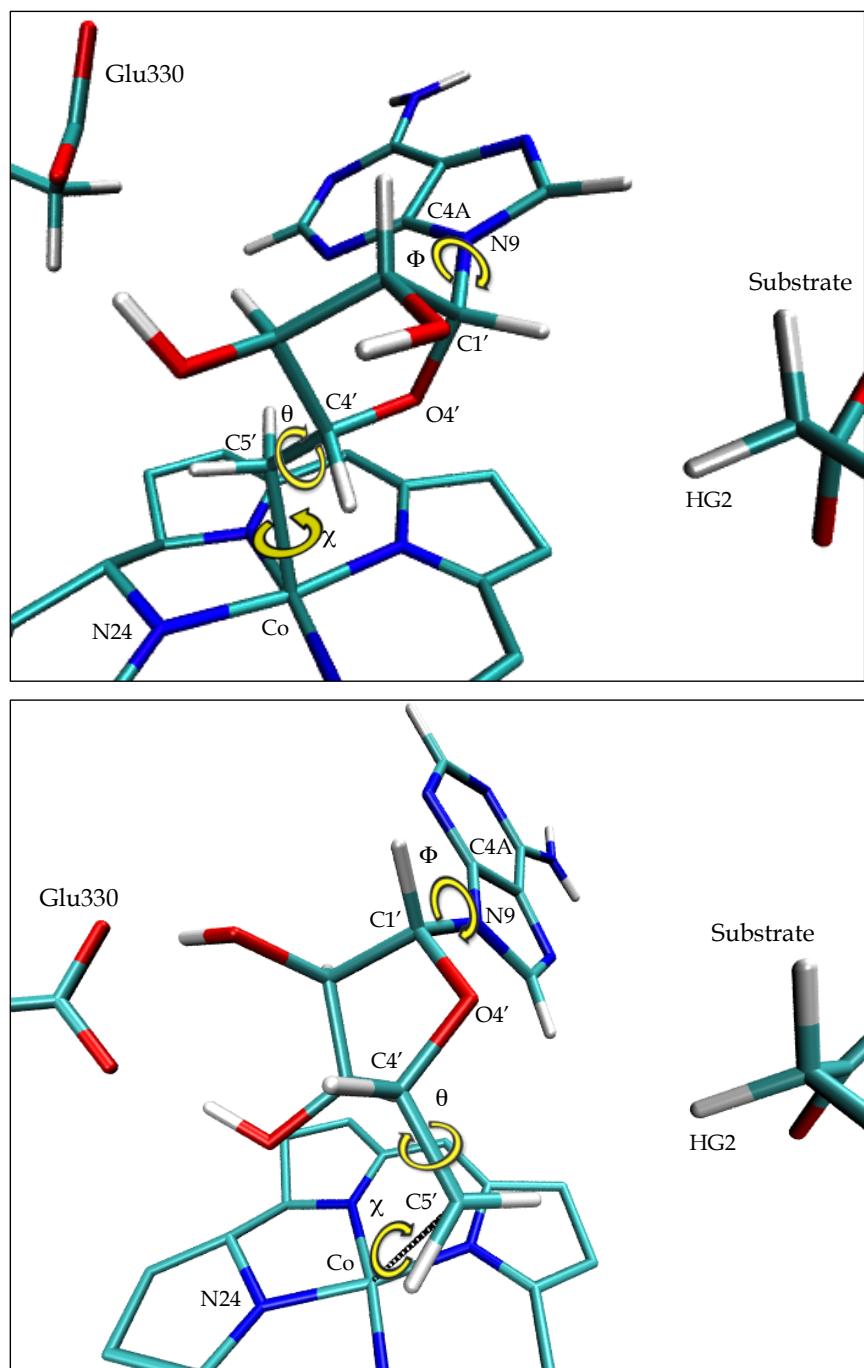


Figure 17 Representative structures of the associated (above) and dissociated (below) states from MD simulations showing the dihedral angles $O4'-Cl'-N9-C4A$ (Φ), $Co-C4'-C5'-O4'$ (θ) and $N24-Co-C5'-C4'$ (χ).

Table 5 Average values and standard deviations in degrees of the dihedral angles shown in Figure 17. The data shown here for each state are the combined data of both active sites and both MD runs (starting from adenosyl conformations A and B). Note that the crystal structure values do differ between the two active sites, although not dramatically. The values for active site between chains A and B are shown here.

a) Data from full simulations

dihedral angle	associated		dissociated		crystal structure	
	av. dihe	SD	av. dihe	SD	conf. A	conf. B
O4'-C1'-N9-C4 (Φ)	79	45	225	11	220	190
Co-C4'-C5'-O4' (θ)	122	84	228	16	239	205
N24-Co-C5'-C4' (χ)	155	46	309	55	302	261

b) Data from last 5ns of each simulation

dihedral angle	associated		dissociated		crystal structure	
	av. dihe	SD	av. dihe	SD	conf. A	conf. B
O4'-C1'-N9-C4 (Φ)	51	12	228	10	220	190
Co-C4'-C5'-O4' (θ)	79	8	229	13	239	205
N24-Co-C5'-C4' (χ)	136	14	316	23	302	261

Figure 20 shows the distributions of the dihedral angles from the last 5 ns of simulation for each system. The distributions of all three dihedral angles in both states for the E330D and E330Q mutants are very similar to their equivalent distributions for the wild type. The E330A and wild type without substrate distributions are also similar to those for the wild type with substrate except that in a few cases the peaks are slightly offset. The distributions for χ and θ in the dissociated simulations of K326M and K326Q span a much greater range of values and are not unimodal. This could indicate that the conformation of the adenosyl radical is much less constrained in these mutants, or it could be that the simulations of the lysine mutants are insufficiently equilibrated. Certainly neither lysine to methionine nor lysine to glutamine are conservative mutations, so it would not be surprising if much longer simulations are required for proper equilibration. Further evidence that the K326M and K326Q simulations may not be equilibrated is that in the associated state simulations of these mutants, the values of Φ are very similar to their dissociated state values. In simulations of all other systems, Φ has one narrow distribution in the associated state centred at around 230° and another narrow distribution in the associated state at around 60°. We anticipate that the values of Φ for these mutants will converge to the same values as others given sufficient further simulation.

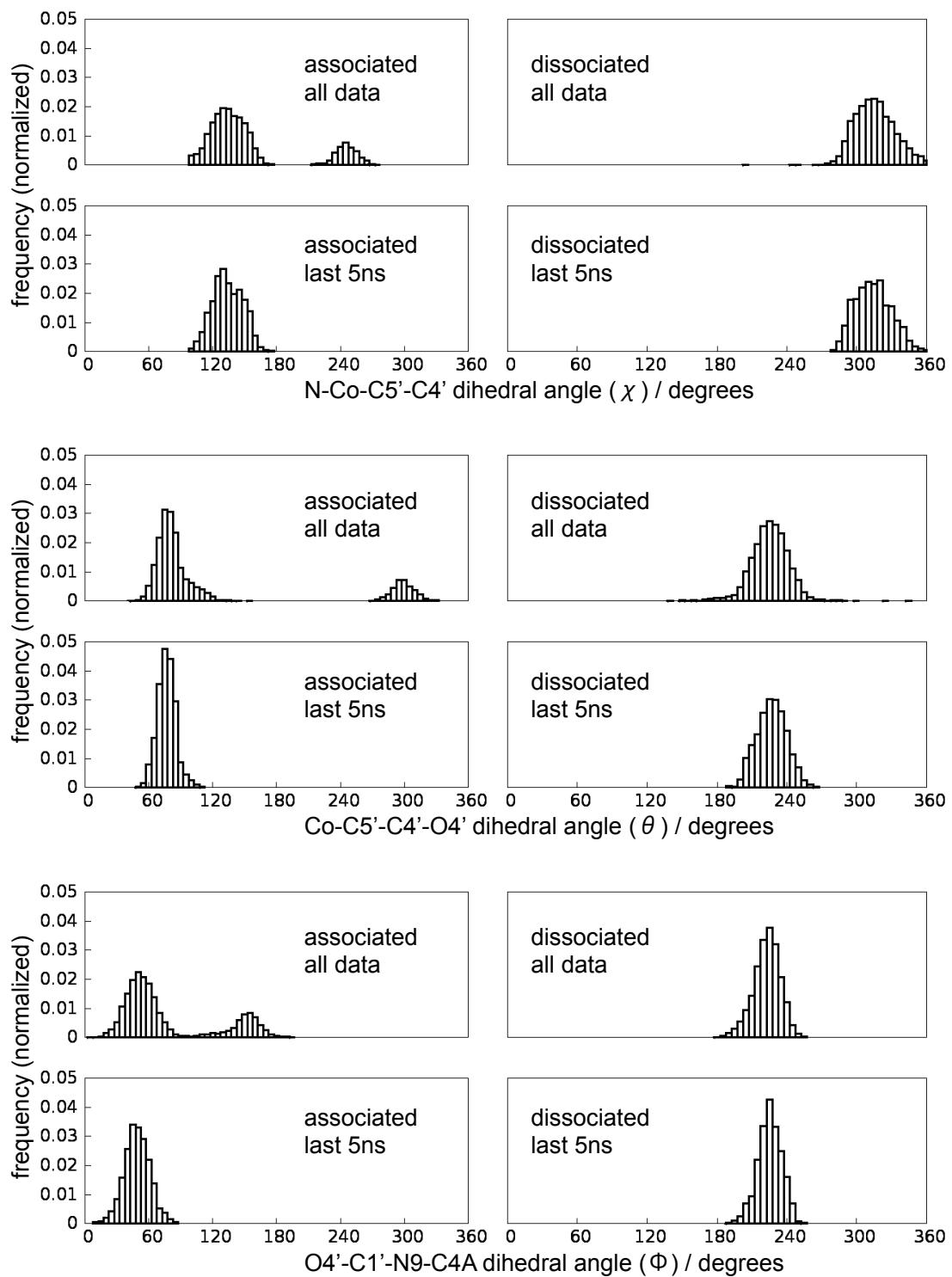


Figure 18 Histograms showing the distributions of the dihedral angles depicted in Figure 17 from whole MD production runs (labelled all data) and from the last 5 ns of each simulation only. The data shown here for each state are the combined data from both active sites and both 20 ns MD runs.

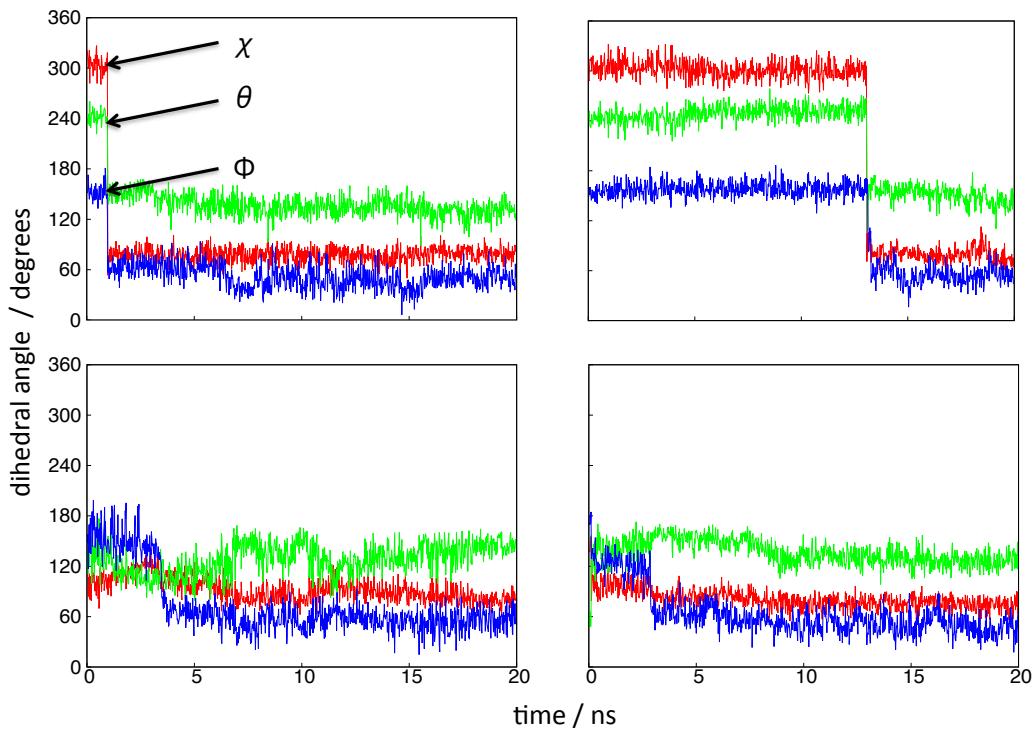


Figure 19 Trace of the dihedral angles shown in Figure 17 over the course of the associated wild type simulations. The top two plots represent the simulations started from adenosyl conformation A, the bottom two from represent those started from conformation B. The plots on the left and right represent data from the two different active sites.

Taken together, the distributions of these dihedral angles from all enzyme systems studied build a fairly consistent picture. The adenosyl moiety adopts one range of internal conformations in the dissociated state and a very different range of conformations in the associated state. These conformations are defined by the average values for the dihedral angles Φ , θ and χ given in Table 5.

One interesting anomaly is the distribution of Φ in the dissociated state of the solution. This has a peak centred at around 55°, whereas the equivalent peaks for all other systems in the dissociated state are centred near 230°. It is inferred that the conformation of Φ adopted in the dissociated state in enzyme must be somehow enforced by the active site structure. Indeed, hydrogen bonding analysis (detailed below) indicates that this appears to be the case. In the dissociated state, the adenine base part of adenosyl is held in place by one set of hydrogen bonds while the ribose –OH groups are held in place by another set. This dictates the value of Φ , which is the dihedral about the bond separating the ribose and adenine groups.

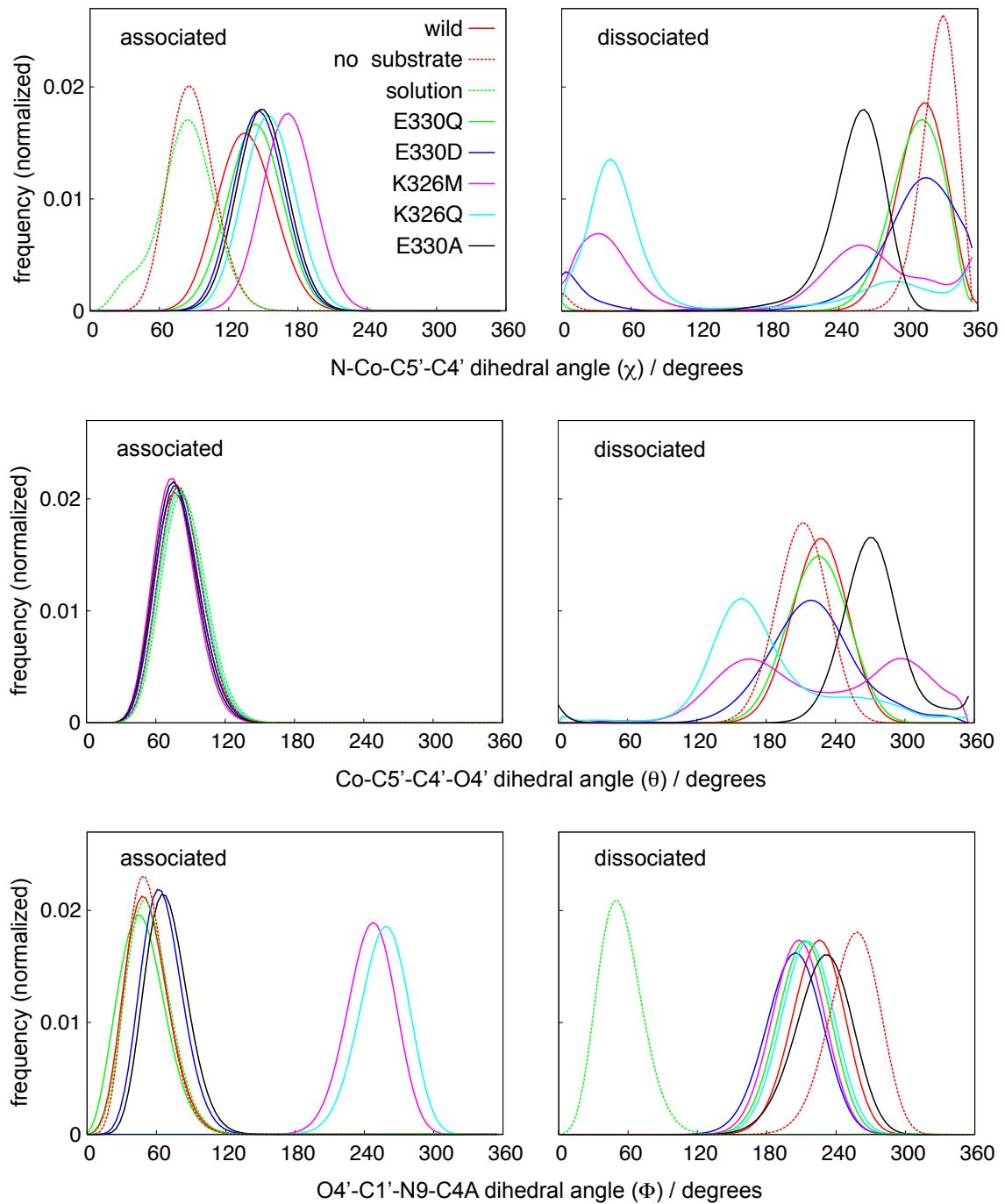


Figure 20 The distributions of the dihedral angles shown in Figure 17 from the last 5 ns only of each MD simulation. The lines shown here are Bézier curves fitted to histograms of the data of the type shown in Figure 18.

4.3.3. Ribose pseudorotation

It has been proposed in the literature that pseudorotation of the adenosyl ribose group controls the hydrogen atom transfer step in adenosylcobalamin dependent enzymes.¹¹⁷ The basis for this is the observed change from C2' endo to C3' endo pseudorotation conformations between the adenosyl crystal structure conformations B and A respectively. It is suggested that this conformational change guides the C5' radical centre on a trajectory towards the substrate. The pseudorotational conformations of the adenosyl ribose group from simulations of the associated and dissociated states in enzyme and solution were analysed to investigate this hypothesis.

One widely used system for conformational analysis of five membered sugar rings was defined by Altona and Sundaralingam.¹²⁸ This system reduces the dimensionality of the conformational space of the ring backbone from five parameters (the dihedral angles of the backbone atoms) to two (the pseudorotational phase angle and the puckering amplitude). The pseudorotational phase angle indicates which atoms are out of plane and whether the atoms in question are cis to the base relative to the plane of the ring (endo) or trans (exo). This is illustrated in Figure 21. The puckering amplitude is a measure of the degree of deviation from planarity of the ring. The distributions of pseudorotational conformations of the adenosyl moiety from wild type and solution simulations are shown in Figure 22.

The dominant conformations for 5 membered sugar rings are known from crystal structures to be C3' endo and C2' endo.¹²⁸ Analysis of the simulations of free adenosyl in solution shows that these were indeed the most highly populated conformations although C4' exo was also highly populated and all of the available conformations were accessible to some extent. The associated adenosylcobalamin in solution was constrained almost entirely to the C3' endo and C4' exo conformations. In contrast, far more conformations were available to the associated state in enzyme (although some conformations remained completely unpopulated). This difference between the associated state in enzyme and solution must be another manifestation of the influence that the active site structure has on the adenosyl conformation. In the dissociated state of the enzyme system, the adenosyl is confined almost entirely to the C3' endo and (to a lesser extent) the C2' exo

conformations. Unlike with the dihedral angles discussed earlier, analysing the last 5 ns only of the simulations does not dramatically change the results qualitatively.

This analysis does partially support the hypothesis that pseudorotation guides the C5' radical centre towards the substrate in that the C3' endo conformation is rigidly enforced in the dissociated state (in enzyme) and that this conformation keeps the C5' close to the substrate (see later). However, our results are not consistent with the theory that the conformation changes from C2' endo to C3', as both of these conformations (as well as others) are accessible in the associated state. Rather it appears that as the Co-C bond is broken, the range of available conformations reduces from many to just C3' endo and C2' exo. These conformations appear to be enforced by the formation of two hydrogen bonds between the adenosyl -OH groups and Glu330 as will now be discussed.

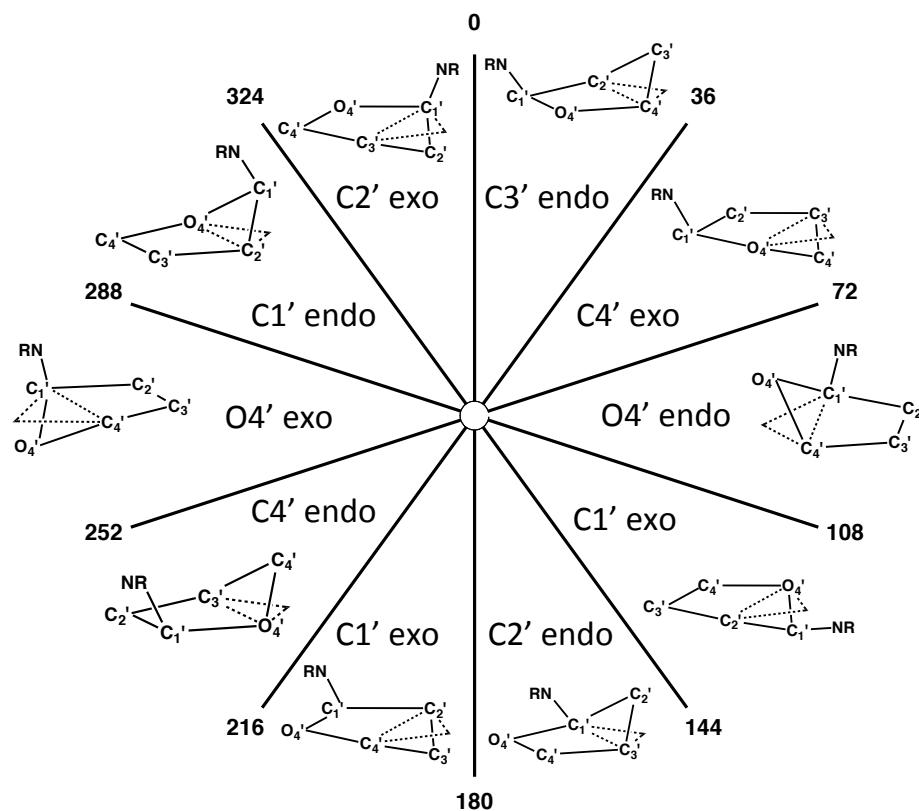
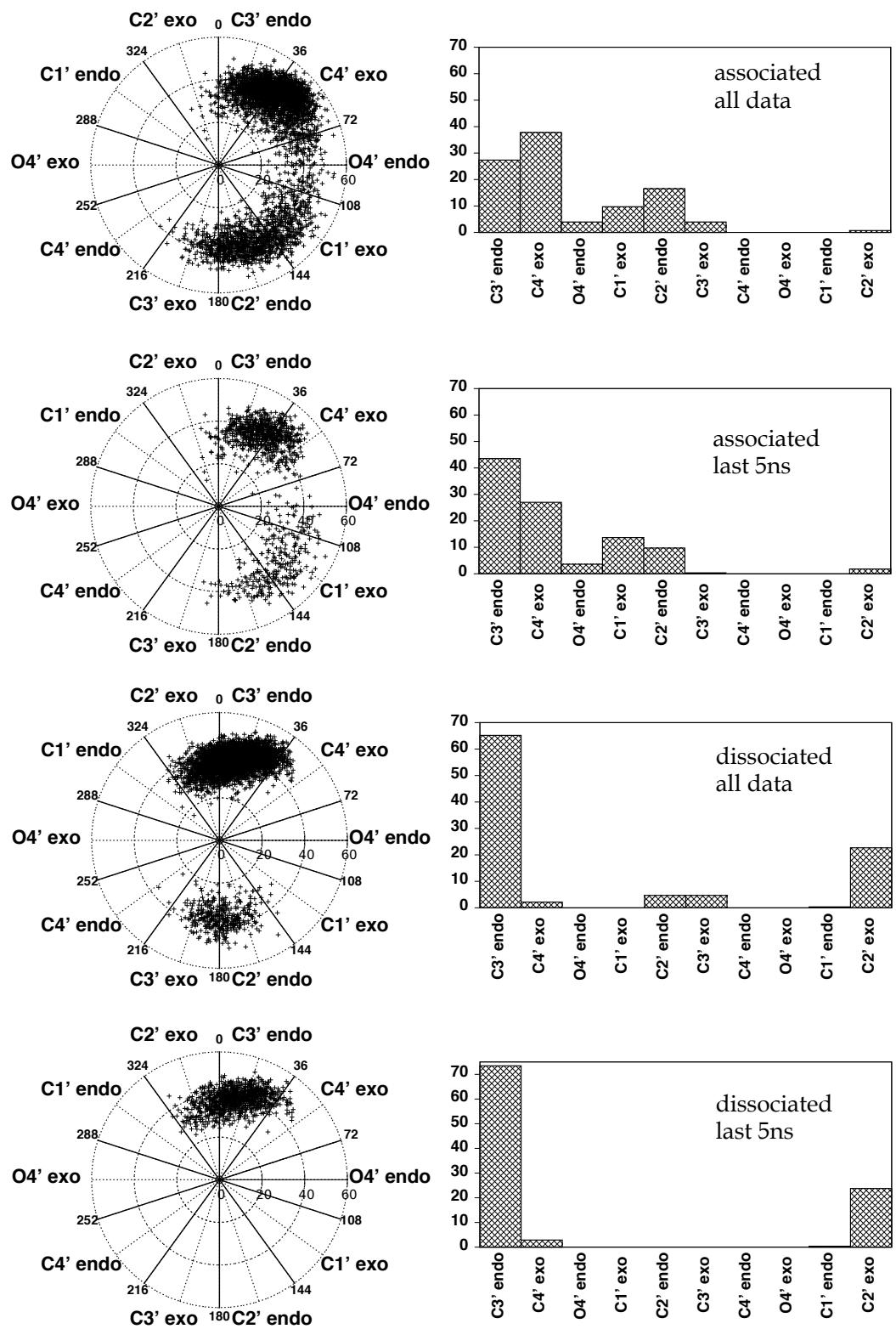


Figure 21 Illustration of the pseudorotaional conformations of five membered sugar rings as defined by the phase angle (polar axis). The convention of Altona and Sundarlingam is used here.¹²⁸



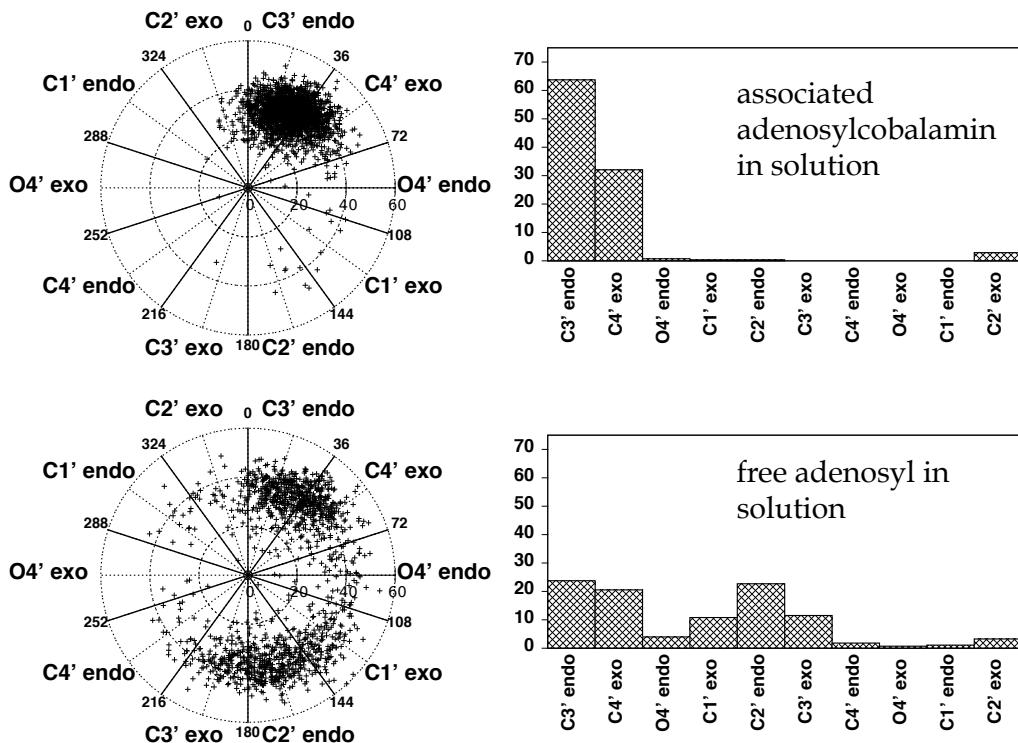


Figure 22 The distribution of pseudorotational conformations of the ribose group of the adenosyl moiety. Each point on the polar plots represents a structure from an MD simulation. The puckering amplitude (radial axis) is plotted against the pseudorotation phase angle (polar axis). The histograms show the percentage of time (y axis) spent in each conformation (x axis) over the course of the simulations. The enzyme data shown here for each state are the combined data from both active sites and both 20ns MD runs. The solution data are from two 20ns MD simulations for each state.

4.3.4. Adenosyl - protein Hydrogen bonding

An EVB study of adenosylcobalamin dependent methylmalonyl-CoA mutase concluded that the entire catalytic effect on Co-C bond breaking was due to electrostatic stabilisation of the dissociated state.¹¹⁹ The study showed that the adenosyl –OH groups act as electrostatic handles which, upon the breaking of the Co-C bond, are free to move closer to and hydrogen bond with the conserved active site glutamate (equivalent to Glu330 in glutamate mutase).

The crystal structure of glutamate mutase indicates that a number of hydrogen bonding interactions exist between the adenosyl moiety and the protein. These are illustrated in Figure 23, in which the hydrogen bonding in the dissociated state closely resembles that of the crystal structure with adenosyl conformation A. The crystal

structure also shows hydrogen bonding interactions to be present between the adenosyl moiety and active site water molecules. These may prove to be important interactions but were not analysed in this work. The adenosyl-protein hydrogen bonds were monitored over the course of the MD trajectories using the hbond facility in the cpptraj model of AmberTools14.¹²⁹ A hydrogen bond was considered formed in any one structure if the donor-acceptor distance was < 3.0 and the donor-hydrogen-acceptor angle was $> 135^\circ$. Analysis of each donor-acceptor pair is summarised in Table 6. To simplify the analysis, the donor-acceptor pairs were collected into three groups: those between the adenosyl – OHs and Glu330 (or mutant equivalents), those between adenosyl –OHs and Lys326 (or mutant equivalents) and those between the adenosyl adenine base –NH₂ group and the protein backbone (the backbone carbonyl oxygens of residues Gly68 and Asn123). This simplified analysis is summarised in Table 7.

The hydrogen bond populations appear to have converged. The values averaged over the last 5 ns of each simulation only are very similar to those averaged over the whole simulations. In agreement with the EVB study of methylmalonyl-CoA mutase, there is considerably more adenosyl-protein hydrogen bonding in the dissociated state than in the associated state. The hydrogen bonds between the adenosyl adenine base and the backbone oxygens from residues Gly68 and Asn123 are highly populated in the dissociated state but are barely populated in the associated state (in fact Gly68 does not hydrogen bond to the adenosyl at all). This appears to be a result of the adenine group being pulled away from the Gly68 and Asn123 as the Co-C bond is formed. In the dissociated state, both of the adenosyl –OH groups form robust hydrogen bonds with Glu330, whereas there is on average less than one of the hydrogen bonds in the associated state. There is a slight but not obviously significant increase in hydrogen bonding between Lys326 and the adenosyl –OH groups in the dissociated state relative to the associated state. The simple counting of hydrogen bonds provides no direct quantitative information on energetics. Crude though this method is, we think it is reasonable to assume that the increased average number of hydrogen bonds in the dissociated state over the associated state will correspond to a significant electrostatic stabilisation of the dissociated state.

The combination of the robust hydrogen bonds between Glu330 and the ribose –OH groups, and between the adenine base –NH₂ group and the protein backbone fixes the positions of the ribose and adenine groups of the adenosyl relative to each other. This results in an enforced value for the dihedral angle Φ of around 230°. This fairly

restrictive hydrogen bonding network in the dissociated state also seems to be responsible for enforcing the other adenosyl dihedral angles θ and χ and the pseudorotational conformation of the ribose group. In the associated state, Φ reverts back to the conformation that it has in the solution simulations of both associated and dissociated states ($\Phi \approx 60^\circ$). This implies that the associated state represents the ideal conformation of Φ , whereas the dissociated state Φ is enforced by the active site structure. The reduced adenosyl-Glu330 hydrogen bonding in the associated state relaxes the constraints on the pseudorotational conformation of the ribose group, resulting in the wider range of conformations observed in the associated state simulations.

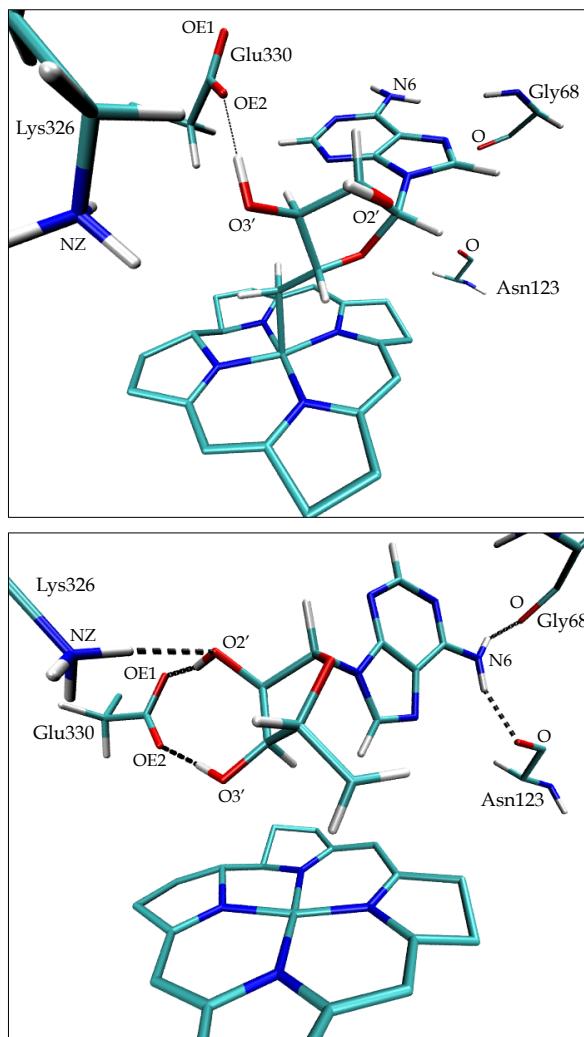


Figure 23 Representative structures of the associated (above) and dissociated (below) states from molecular dynamics simulations showing hydrogen bonding interactions between the adenosyl moiety and the protein.

Table 6 Percentage formed, average distances and standard deviations (in Å) of the hydrogen bonds shown in Figure 23. Data taken from full simulations of the wild type enzyme in the associated and dissociated states.

acceptor	donor	associated			dissociated			crystal structure	
		%	av. dist	SD	%	av. dist	SD	conf. A	conf. B
Glu330 OE1	Ado O2'	4	4.9	1.2	75	2.8	0.3	2.5	3.2
Glu330 OE2	Ado O2'	5	4.4	1.1	22	3.1	0.5	3.3	4.4
Glu330 OE1	Ado O3'	20	4.5	1.3	24	3.6	0.6	3.9	3.0
Glu330 OE2	Ado O3'	33	3.9	1.1	75	2.8	0.4	2.7	3.4
Ado O2'	Lys326 NZ	10	4.1	0.9	37	3.3	0.8	3.8	5.4
Ado O3'	Lys326 NZ	12	3.3	0.4	6	3.3	0.3	3.3	2.8
Glu330 OE1	Lys326 NZ	16	4.4	1.0	8	3.6	0.5	4.0	4.0
Glu330 OE2	Lys326 NZ	24	3.9	0.9	10	3.2	0.5	3.2	3.2
Gly68 O	Ado N6	0	5.2	1.0	44	3.0	0.2	3.2	2.9
Asn123 O	Ado N6	11	5.1	1.7	69	2.9	0.2	2.7	3.1

Table 7 Average number of hydrogen bonds between the adenosyl moiety and the protein, arranged into three groups of hydrogen bonds: those between the adenosyl -OHs and Glu330, those between the adenyl -OHs and Lys326 and those between the -NH₂ of the adenine base of the adenosyl moiety and the Gly68 and Asn123 backbone oxygens.^a

a) Data from full simulations

	330-Ado	326-Ado	base-backbone	total	diss. - ass.
associated	0.62	0.22	0.11	0.94	
dissociated	1.96	0.43	1.12	3.51	2.56

b) Data from last 5ns of each simulation

	330-Ado	326-Ado	base-backbone	total	diss. - ass.
associated	0.60	0.17	0.06	0.84	
dissociated	2.03	0.41	1.07	3.52	2.68

Table 8 summarises the hydrogen bonding analysis for wild type, wild type with no substrate and all mutant enzymes. It has been proposed in the literature that substrate binding may induce some active site conformational change that facilitates Co-C bond breaking.¹²⁰ In the simulations of the wild type with no substrate in the dissociated state, there is much less hydrogen bonding between the adenine base -NH₂ group and the protein backbone than in the equivalent simulations with substrate. This results in a smaller difference in number of hydrogen bonds between the two states. It appears that

^a The reader may wonder why it is possible for an average of greater than two hydrogen bonds to be formed between Glu330 and the adenosyl -OH groups. This is due to the relatively crude definition of a hydrogen bond used here, which counts hydrogen bonds between one -OH group and both carboxyl oxygens of Glu330 in certain conformations.

the presence of substrate may be required to preorganise properly the active site for maximal electrostatic stabilisation of the dissociated state.

The residues Glu330 and Lys326 were chosen for mutation to investigate the importance of hydrogen bonding between these residues and the adenosyl –OH groups. The mutants exhibited significant but varying reductions in activity relative to the wild type (Table 9). We will now examine the hydrogen bonding populations from the MD simulations in an attempt to rationalise the observed experimental kinetics in terms of changes in proposed electrostatic stabilisation of the dissociated state.

The wild type (with substrate) does have the most adenosyl-protein hydrogen bonding in the dissociated state of any of the systems studied. However, it also has the most hydrogen bonding in the associated state. Consequently the greatest difference in hydrogen bonding between the two states is not observed in the wild type. Rather it is the slowest mutant, K326Q, which exhibits the greatest difference and therefore (according to our crude approximation) should provide the greatest electrostatic stabilisation of the dissociated state. E330A exhibits the least adenosyl-protein hydrogen bonding in both states as well as the lowest difference in hydrogen bonding between the two states. While this mutant does have very low activity, it does not have the lowest as would be predicted from the hydrogen bonding analysis alone. E330D exhibits almost exactly the same hydrogen bonding behaviour as the wild type (which is not surprising as the only difference introduced by this mutation is the shortening of the sidechain by one –CH₂ group), yet it is much less active than the wild type and is not even the most active mutant.

Further interpretation of the hydrogen bonding patterns is unlikely to provide much insight into the relative mutant activities, as it is clear already that they cannot be rationalised by this analysis alone. Despite the crudeness of relating the number of hydrogen bonds to the electrostatic stabilisation of the dissociated state, these results suggest that there are other factors influencing the relative rates of tritium exchange. As it is known that Co-C bond breaking and hydrogen abstraction are kinetically coupled in glutamate mutase, the difference in mutant activities may be related to perturbations to the rate of hydrogen abstraction as well as Co-C bond breaking.

Table 8 Average number of hydrogen bonds between the adenosyl moiety and the protein (as in Table 7) for the wild type with substrate, wild type with no substrate present and for the five mutant systems studied. Data taken from both active sites and last 5 ns of each simulation.

	330 - Ado	326-Ado	base-backbone	total	diss. - ass.
wild type					
associated	0.62	0.22	0.11	0.94	
dissociated	1.96	0.43	1.12	3.51	2.56
no substrate					
associated	0.09	0.49	0.00	0.59	
dissociated	1.81	0.51	0.17	2.49	1.90
E330A					
associated	0.00	0.16	0.00	0.16	
dissociated	0.00	0.10	1.16	1.26	1.10
E330D					
associated	0.15	0.24	0.10	0.49	
dissociated	1.64	0.12	1.37	3.13	2.65
E330Q					
associated	0.12	0.14	0.23	0.49	
dissociated	1.05	0.44	1.06	2.54	2.05
K326M					
associated	0.38	0.00	0.21	0.60	
dissociated	1.57	0.00	1.30	2.86	2.27
K326Q					
associated	0.07	0.01	0.11	0.19	
dissociated	1.57	0.02	1.38	2.97	2.78

4.3.5. Positioning of the radical centre on C5' carbon for hydrogen abstraction

How reactions are facilitated is one of the key questions in the study of any catalytic system. In adenosylcobalamin dependent enzymes and other systems exploiting highly reactive intermediates, an equally important question is how these intermediates are controlled so as to prevent unwanted side reactions. The ability of enzymes to do this has been dubbed negative catalysis.¹³⁰ As already discussed, it has been proposed in the literature that glutamate mutase employs ribose pseudorotation as a means to carefully direct the C5' radical centre towards the substrate. So far in this chapter a picture has begun to emerge of an orchestrated set of conformational changes accompanying Co-C bond breaking. In the following, the effect that the hydrogen bonding induced restrictions on the adenosyl conformation have on the orientation of the C5' radical centre will be investigated.

After formation by Co-C bond cleavage, the adenosyl radical abstracts a hydrogen atom from the CG carbon of the glutamate substrate. The distance between the C5' carbon radical centre and the HG2 hydrogen during the MD simulations was recorded, as was the C5'-HG2-CG angle. These are illustrated in Figure 24. For the purposes of this analysis it was assumed that the hydrogen abstracted is the HG2 hydrogen. This is because during the MD simulations, HG2 is usually closer than HG3 to the C5' carbon. The distributions of these parameters are shown in Figure 25 and Figure 26. The average values of the C5'-HG2 distance for the wild type and mutants are given in Table 9 as are the experimentally determined rate constants.

Clearly the C5' carbon must come into close contact with the relevant hydrogen in order for abstraction to occur. For reference, the values for the C5'-H distance and the C5'-H-CG angle at the transition structure of the hydrogen transfer of the model system described in section 3.4 were 1.4 Å and 173° respectively. The distribution of C5'-H distances for the wild type is fairly narrow and centred at a relatively short distance. This is consistent with the hypothesis that the active site of glutamate mutase is preorganised so as to position optimally the C5' radical centre for hydrogen abstraction. The distribution of angles is fairly broad. Although the model system transition structure angle is not highly populated in the MD simulations, it is clearly accessible.

The mutant systems all exhibit wider C5'-H distance distributions centred at longer distances than the wild type. The E330A mutant has the broadest distribution centred furthest to the right. This is consistent with the fact that there are no hydrogen bonds between this residue and the adenosyl -OH groups and therefore fewer restraints on the adenosyl position in this mutant than in the other systems.

The C5'-H distance tends to be significantly longer in E330D than in the wild type. The aspartate sidechain is shorter than that of glutamate. As it is covalently bonded to adjacent amino acid residues, the aspartate is not expected to be able to move as much as the adenosyl radical. Neither is the substrate as it is tightly held in place by several salt bridges. As the adenosyl -OH groups hydrogen bond with the aspartate carboxylate group, the net effect is that the whole adenosyl group is pulled further away from the substrate, explaining the longer C5'-H distances relative to the wild type.

The correlation between the average C5'-HG2 distances from MD and experimentally determined tritium exchange rates constant is shown in Figure 27. There does appear to be a significant correlation between the average C5'-H distance of a particular system and activity. In general the less active mutants appear less able to

position optimally the C5' carbon for hydrogen abstraction. The correlation is not perfect and there clearly is variation in the hydrogen bonding patterns of the mutants. It appears that electrostatic stabilisation of the dissociated state and optimal C5' radical positioning are both consequences of an active site preorganised to bind the adenosyl radical in a specific conformation via numerous hydrogen bonds. It is proposed that both of these factors are important to catalysis and that the observed differences in mutant activity originate from perturbations to both of them. The methods for quantifying these factors used in this chapter are approximate. As such it is not considered worthwhile to attempt to combine them to reproduce or rationalise the experimental results. In the following chapter, however, the barriers to hydrogen abstraction for the wild type and mutant systems will be explicitly calculated.

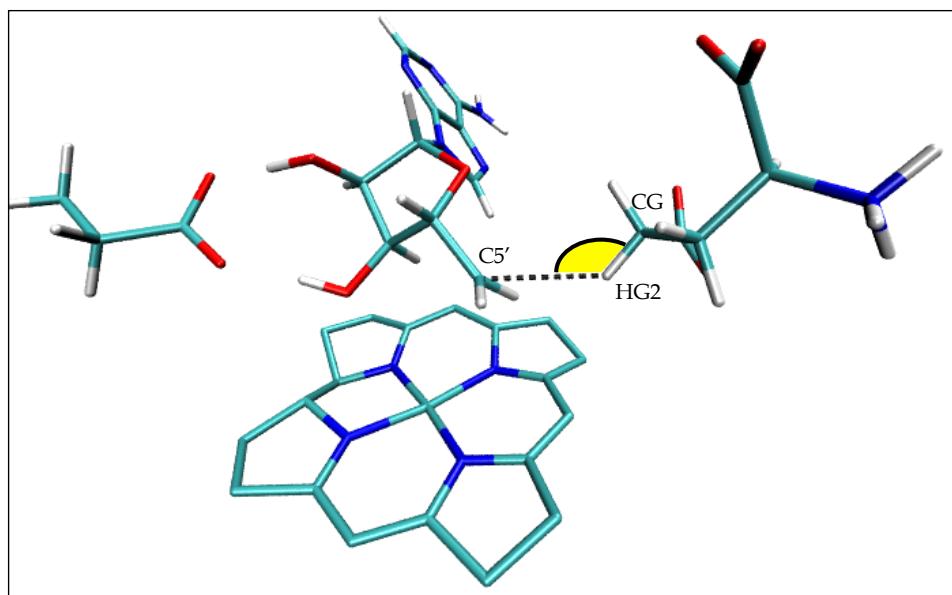


Figure 24 Illustration of the distance between the carbon radical center of the adenosyl radical and the abstractable hydrogen of the glutamate substrate R(C5'-HG2). Also illustrated by the yellow arc is the angle A(C5'-HG2-CG).

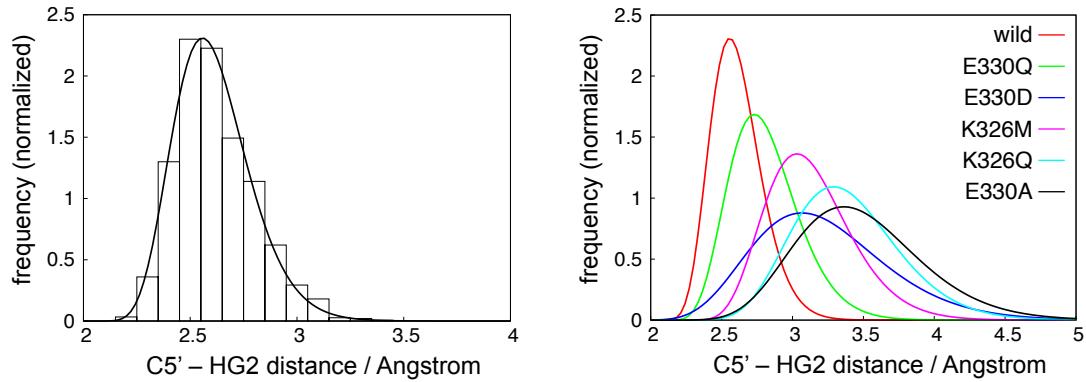


Figure 25 Distributions of the distance $R(C5'-HG2)$ shown in Figure 24. The lines are Rayleigh functions fitted to histograms of the data as shown on the left for the wild type. The image on the right shows the distributions for the wild type and mutant systems.

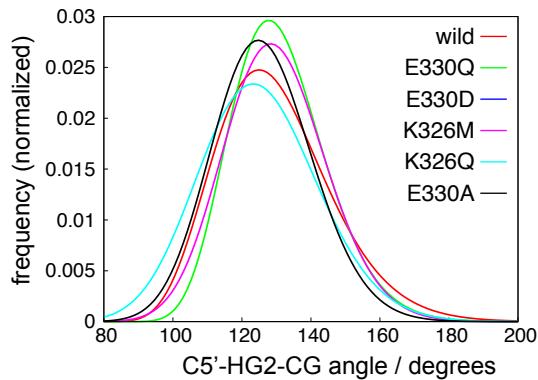


Figure 26 Distributions of the angle $A(C5'-HG2-CG)$ shown in Figure 24. The lines are Rayleigh functions fitted to histograms of the data.

Table 9 Comparison of average C5'-HG2 distances from MD simulations with rate constants derived from experiments conducted by our collaborators at the University of Michigan.¹¹⁸

	$k_T \times 10^3 \text{ s}^{-1}$	$\log(k_T)$ relative to		$R(C5'\text{-H}) / \text{Angstroms}$
		wild type		
wild	5000 ± 100	0		2.67 ± 0.18
E330Q	50 ± 2	-2.0		2.85 ± 0.23
E330D	20 ± 1	-2.4		3.27 ± 0.44
K326M	8 ± 0.1	-2.8		3.3 ± 0.62
E330A	0.9 ± 0.04	-3.7		3.61 ± 0.56
K326Q	0.7 ± 0.02	-3.9		3.43 ± 0.39

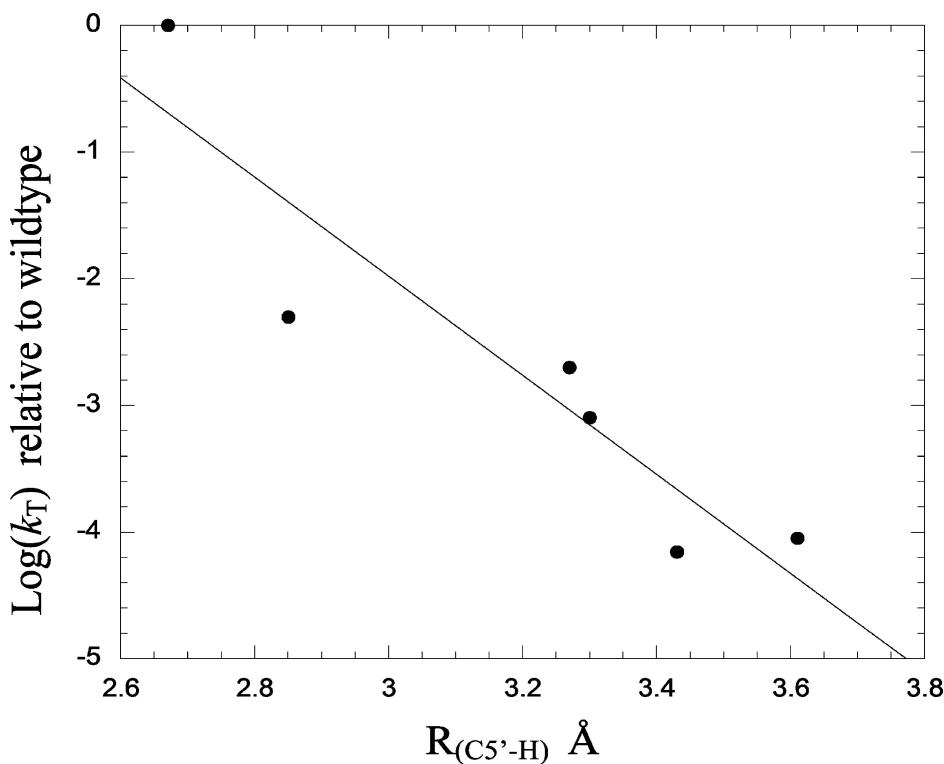


Figure 27 Logarithms of the experimental rate constants for the wild type and mutant enzymes plotted against the average C5'-HG2 distances from MD simulations. Figure taken from a paper published by the author and collaborators.¹¹⁸

Chapter 5. Modelling Co-C bond breaking and hydrogen abstraction

5.1. Introduction

Despite decades of research into adenosylcobalamin dependent enzymes from numerous computational and experimental groups, no consensus has been reached as to exactly how Co-C bond breaking is catalysed. The nature of the coupling between the Co-C bond breaking and subsequent hydrogen abstraction also remains open for debate.⁷⁴ Some claim that these steps are concerted¹¹⁹ while others propose a stepwise mechanism.¹³¹ Certainly the adenosyl radical, if it exists as an intermediate, must have a short lifespan as it has never been detected directly by experimental techniques such as EPR spectroscopy.¹³² In contrast, various intermediates from the radical rearrangement that follows hydrogen abstraction have been observed to accumulate during turnover of glutamate by glutamate mutase. This has led to the conclusion that no one elementary step is cleanly rate limiting and that the equilibrium constant for Co-C homolysis is close to 1.^{132,133} The steps following hydrogen abstraction are clearly important to catalysis and worthy of study. However, these steps are now thought to be well understood and it is assumed that their modelling can be neglected for the purposes of this work focusing on Co-C bond breaking and hydrogen abstraction in isolation.^{72,76,134–136}

Different computational studies have produced qualitatively very different reaction profiles for Co-C bond breaking in enzyme.^{33,119,131,137} It is clear from the results of the previous chapter that significant conformational changes accompany this step. It is proposed that a major cause for the lack of consistency between the previous studies is their inability to capture properly these effects (as a result of their limited conformational sampling). In order to check for convergence and internal consistency of the calculated reaction profiles, Co-C bond breaking was modelled in the forward (bond breaking) and backward (bond forming) directions starting from a range of structures taken from MD simulations of the associated and dissociated states respectively. Although some insight

into Co-C bond breaking was gained from the modelling performed in this work, internal consistency and convergence were not achieved. Because of this, calculations of this step were not attempted for the mutant systems discussed in the previous chapter. The results for the hydrogen abstraction step were more robust, so this step was modelled for the wild type and mutant systems in an attempt to reproduce and rationalise the observed differences in activity.

5.2. Computational details

5.2.1. Starting structures

Starting structures for modelling Co-C bond breaking in the forward and backward directions were taken from MD simulations of the associated and dissociated states respectively (as described in section 4.2.2). The hydrogen abstraction step was modelled only in the forward direction starting from structures taken from MD simulations of the dissociated state. Starting structures were extracted at intervals of 1 ns from the MD trajectories. Both active sites were used for modelling the Co-C bond breaking with adiabatic mapping on the EVB potential. Active site A only was used for all other calculations presented in this section.

5.2.2. Adiabatic mapping with EVB and QM/MM

Before adiabatic mapping on the EVB potentials, each structure was truncated by deleting all residues with no atoms within 15 Å of the adenosyl moiety (of the active site being studied). All atoms further than 10 Å from the adenosyl moiety were frozen during the minimisations.

For adiabatic mapping of the Co-C bond breaking with both EVB and QM/MM, a harmonic restraint of 2000 kcal/mol/Å² was placed on the distance R(Co-C). The minimum of this harmonic restraint was increased in steps of 0.1 Å from 1.8 to 3.5 Å for adiabatic mapping in the forward direction and decreased in steps of 0.1 Å from 3.5 to 1.8 Å for adiabatic mapping in the backward direction.

For adiabatic mapping of the hydrogen abstraction step, with both EVB and QM/MM, a harmonic restraint was placed on the reaction coordinate R(CG-HG2) – R(C5'-HG2). The minimum of this harmonic restraint was increased from -2.0 Å to 2.0 Å

in steps of 0.2 Å. For the purposes of this study, we have assumed that it is the HG2 hydrogen that is abstracted by the adenosyl radical. This assumption is not rigorously justified and was based solely on the fact that the C5' carbon radical centre is generally closer to the HG2 hydrogen than the HG3 hydrogen in the unbiased MD simulations of the dissociated state (this is also the case in the crystal structure). In principle either hydrogen could be abstracted to give an intermediate that is chemically identical.

Before adiabatic mapping with QM/MM, each structure extracted from the MD simulations was minimised using the sander module of Amber11.¹²⁶ The system was then truncated. Chains C and D and their associated adenosylcobalamin and substrate groups were removed, as were all water molecules and ions with no atoms within 3 Å of chains A or B. The amber prmtop and rst files were then converted into a Tinker xyz file using in-house scripts which call the Tinker module pdb2xyz.¹³⁸

The QM/MM calculations were performed with QoMMA which called Jaguar version 7.8 for all QM and Tinker Version 6.1 for all MM calculations.^{113,138,139} Unrestricted DFT at the B3LYP*/LACVP* level of theory was used for the QM components of the calculations. The amber99sb force field with extra parameters for adenosylcobalamin (as described in section 3.2) was used for the all MM components.

For the QM/MM calculations of Co-C bond breaking in glutamate mutase, the QM region was defined as the substrate, the adenosyl moiety, the imidazole group of the Co coordinating histidine and a truncated corrin ring as shown in Figure 8. For calculations of the hydrogen abstraction, the QM region consisted of the adenosyl moiety and the substrate only. In each case the active region (in which atoms are allowed to move during minimisations) was defined as containing all residues with at least one atom within 6 Å of the QM region. All atoms outside this region were frozen.

The dissociated state of adenosylcobalamin is a biradical open shell singlet. Without a broken symmetry open shell singlet initial guess, the wavefunction will usually converge to a closed shell singlet. The 2spin and fragment commands in Jaguar were used in order to generate the necessary initial guesses. For the QM/MM adiabatic mapping in the backward direction a broken symmetry initial guess was simply generated and used for the starting structure. When performing adiabatic mapping in the forward direction a different procedure was required, as the associated state is a closed shell singlet. The Co-C bond was stretched by adiabatic mapping from 2 Å to 2.5 Å, at which point a broken symmetry open shell singlet wavefunction was generated. Using the resulting

wavefunction and structure, adiabatic mapping was then performed in the forward direction from 2.5 to 3.6 Å and in the backward direction from 2.5 to 1.8 Å.

Any potential energy profiles from adiabatic mapping that had discontinuities in them were discarded.

5.2.3. Umbrella sampling system setup

For the umbrella sampling simulations in enzyme stochastic boundary conditions were used.¹⁴⁰ A similar protocol to that described in previous work by our group was used to set up and run the stochastic boundary MD simulations.¹⁴¹ All atoms more than 31 Å from the adenosyl C5' carbon were deleted. The reservoir region was defined as containing all atoms further than 25 Å from the C5' carbon. Atoms in the reservoir region were held fixed. The reaction region (in which standard Newtonian MD is performed) consisted of all atoms less than 20 Å from the C5' carbon. The buffer region (in which Langevin dynamics is performed) consisted of all atoms between 20 and 25 Å from the C5' carbon. Friction coefficients of 62 ps⁻¹ and 250 ps⁻¹ were applied to heavy atoms belonging to the buffer region protein and water molecules respectively. Harmonic restraints were put on all heavy protein atoms in the buffer region. The force constants for these restraints increased incrementally from the inside of the buffer region to the outside and were originally derived from x-ray temperature factors of a different enzyme system.¹⁴² A deformable boundary potential of radius 25 Å and centred on the C5' carbon was applied to all water molecules.¹⁴⁰

A bug was found in the CHARMM code that prevented the correct application of stochastic boundary conditions when combined with EVB (or any CHARMM module that requires CHARMM to run in parallel). This bug was fixed in our local version of CHARMM but remains in the main release of the code.

Periodic boundary conditions were used for all umbrella sampling simulations in solution (TIP3P water). For the Co-C bond breaking simulations in solution a cubic periodic box of length 35 Å was used. The ribonucleotide tail of adenosylcobalamin was removed and an imidazole group was coordinated to the Co. Apart from these modifications, the full adenosylcobalamin molecule was modelled.

For the umbrella sampling simulations of the hydrogen abstraction step in solution, only the adenosyl and substrate moieties were modelled. A cubic box of length 25 Å was used.

5.2.4. Umbrella sampling procedures

The SHAKE algorithm was used to constrain the rigidity of all TIP3P water molecules but was not used for any other bonds involving hydrogen. An integration timestep of 1 fs was used for all simulations.

Firstly, umbrella sampling simulations were performed of the Co-C bond breaking in the forward and backward directions with the R(Co-C) distance as the only reaction coordinate. A harmonic restraint of 200 kcal/mol/Å² was placed along R(Co-C), with equilibrium values of 2 Å and 3.8 Å for the simulations in the forward and backward direction respectively. With the stochastic boundary conditions in place, the system was then minimised for 3000 steps of adopted basis Newton-Raphson. Then the system was heated from 100 to 300 K in increments of 5 K every 0.1 ps for 20 ps followed by 100 ps of MD. Then, MD of the first umbrella sampling window was performed. At each window 50 ps of equilibration MD was performed followed by 400 ps of production MD (during which the value of the reaction coordinate was recorded every timestep). The equilibrium value of the reaction coordinate restraint was then increased (or decreased in the case of the simulations in the backwards direction) by 0.1 Å; ready for simulation of the next window. This was repeated until the equilibrium value of the restraint reached 3.8 Å in the forward direction or 1.8 Å in the backward direction.

The weighted histogram analysis method (WHAM) as implemented in the code of the Grossfield lab was used to generate the potential of mean force (PMF) free energy profiles from all of the umbrella sampling simulations reported here.⁵⁸ For each simulation, histograms of the distribution of values of the reaction coordinates for each window were plotted. These plots were inspected to ensure that there was sufficient overlap between each window. This is essential in order for WHAM to produce meaningful free energy profiles.

Umbrella sampling of the hydrogen abstraction step was performed using a very similar procedure. A harmonic restraint of 200 kcal/mol/Å² was placed along the coordinate R(CG-HG2) – R(C5'-HG2). The equilibrium value of this restraint was increased from -2 Å to 2 Å in steps of 0.2 Å.

Umbrella sampling of the dihedral angle D(O4'-C1'-N9-C4A) was done following a similar procedure. For sampling of the associated and dissociated states, the appropriate single state (non-EVB) force field was used as described in section 3.2. A harmonic restraint of 65.7 kcal/mol/radian² was imposed on the reaction coordinate. In steps of 10°,

the minimum of this restraint was increased from 0° to 360° for the associated state simulations and decreased from 200° to -165° for the dissociated state simulations.

5.2.5. Two-dimensional umbrella sampling

The free energy profiles of Co-C bond breaking in enzyme did not converge with respect to the direction of simulation using a single reaction coordinate (as will be discussed later). In an attempt to address this issue, two-dimensional umbrella sampling was performed. In each case the distance R(Co-C) was used as one coordinate. The second coordinate was either the dihedral angle D(Co-C5'-C4'-O4') or the dihedral angle D(Co-C5'-C4'-O4'), as described in section 4.3.2. For these simulations a harmonic restraint of 200 kcal/mol/Å² was placed on the coordinate R(Co-C) and a restraint of 65.7 kcal/mol/radian² was placed on the dihedral angles.

For the first umbrella sampling window the equilibrium values for these restraints were placed in the approximate vicinity of the average value of the second reaction coordinate taken from MD simulations of the relevant state (see section 4.3.2). For the simulations of D(O4'-C1'-N9-C4A) vs R(Co-C) the starting values were D(O4'-C1'-N9-C4A) = 20°, R(Co-C) = 1.8 Å for the associated state and D(O4'-C1'-N9-C4A) = 260°, R(Co-C) = 4.0 Å for the dissociated state. For the simulations of D(Co-C4'-C5'-O4') vs R(Co-C) the starting values were D(Co-C4'-C5'-O4') = 75°, R(Co-C) = 1.8 Å for the associated state and D(Co-C4'-C5'-O4') = 235°, R(Co-C) = 4.0 Å for the dissociated state.

As with the one-dimensional umbrella sampling of R(Co-C), minimisation, heating and equilibration was performed. Then, for each umbrella sampling window, 20 ps of equilibration followed by 50 ps of production (where the values of both reaction coordinates were recorded every 100 steps) was performed. Along the main simulation branch, the equilibrium value of the restraint on the R(Co-C) coordinate was increased by 0.1 Å after each window and the final structure and velocities were used to start the MD simulation in the next window. At the end of the simulation of each window on the main branch, a new branch of simulations was launched in which the R(Co-C) coordinate was restrained at the value of the current window and the second reaction coordinate was increased or decreased between each window of the new branch. For the simulations of D(O4'-C1'-N9-C4A) vs R(Co-C), the secondary branches ran in steps of 10° from D(O4'-C1'-N9-C4A) = 20° to 260° and D(O4'-C1'-N9-C4A) = 260° to 20° in

simulations of the forward and backward directions respectively. For the simulations of D(Co-C4'-C5'-O4') vs R(Co-C), two branches were launched at the end of each main branch window. For the simulations of the forward direction, one new branch increased, in steps of 20° from D(Co-C4'-C5'-O4') = 75° to 295°, and the another branch decreased from D(Co-C4'-C5'-O4') = 75° to -185° (-185° = 175° due to the 360° periodicity of dihedral angles). For the simulations of the backward direction, one new branch increased in steps of 20° from D(Co-C4'-C5'-O4') = 235° to 435° and the another branch decreased from D(Co-C4'-C5'-O4') = 235° to -45°.

2D WHAM analysis was performed with periodicity of 360° along the dihedral coordinates. Histograms of the values of the R(Co-C) reaction coordinate from simulations in windows with the same value of the second reaction coordinate were plotted together to check for overlap between adjacent windows. This was also done for histograms of the values of the second reaction coordinate from windows with the same value of R(Co-C).

Hydrogen bonding analysis was performed in AmberTools as described in section 4.3.4. Analysis of dihedral angles was performed within CHARMM.

5.3. Results and discussion

5.3.1. Adiabatic mapping of Co-C bond breaking

Figure 28 shows the potential energy profiles for the Co-C bond breaking step generated by adiabatic mapping with EVB and QM/MM. The profiles generated by scanning in the forward direction with EVB (top left) can clearly be split into two clear groups. Those shown in red all have a shallow minimum at around 2.6 Å, while those in black exhibit no minima (other than that associated with the formed Co-C bond) and are all very similar to the gas phase Co-C bond breaking profile shown in Figure 7. The starting structures of all of the profiles were inspected. It was found that all of the profiles exhibiting an additional minimum were generated starting from structures taken from early on in an MD simulation of the associated state with the adenosyl group in crystallographic conformation A. In all of these structures, the dihedral angles Φ , θ and χ were far from their equilibrium values for the associated. As discussed in 4.3.2, these structures are not representative of the associated state at equilibrium; rather they are

artefacts of the relatively long time required for the system to relax from the crystal structure with the adenosyl in conformation A (which closely resembles the dissociated state) to an equilibrium conformation of the associated state. The profiles shown in black were all generated from starting structures that were much more representative of the equilibrium associated state, as indicated by their values of Φ , θ and χ .

The profiles generated by adiabatic mapping with EVB in the backward direction exhibit a wide range of characteristics. At one extreme, some exhibit a highly exothermic Co-C bond breaking profile with a very small barrier. At the other extreme, some exhibit an endothermic profile with a barrier of around 10 kcal/mol. The other profiles are spread evenly between these extremes, some of which closely resemble those generated by mapping the forward direction shown in red.

Although the profiles for the Co-C bond breaking in the forward direction calculated with QM/MM (bottom left) rise less steeply in energy than their EVB counterparts, they exhibit qualitatively the same behaviour. The QM/MM profiles generated by adiabatic mapping in the backward direction lie within the range of those calculated with EVB.

It is clear that these potential energy profiles have not converged even qualitatively with respect to the direction of adiabatic mapping. In an attempt to understand why this is, the active site structures from each EVB profile at $R(\text{Co-C}) = 2.0 \text{ \AA}$ and 3.5 \AA were examined and compared.

The total number of adenosyl - enzyme hydrogen bonds was counted for these structures, as described in section 4.3.4. Figure 29 shows the change in number of hydrogen bonds for each profile in going from $R(\text{Co-C}) = 2.0 \text{ \AA}$ to 3.5 \AA or the reverse. The expected numbers of hydrogen bonds for the associated and dissociated states are 0.84 and 3.52 respectively based on averages from MD. In no case in either the forward or backward directions did the number of hydrogen bonds change between 1 in the associated state, and 3 or 4 in the dissociated state. In the cases of adiabatic mapping in the forward direction, some pathways, starting with either 2 or 3 hydrogen bonds, did result in 3 hydrogen bonds by $R(\text{Co-C}) = 3.5 \text{ \AA}$. In none of the calculations in the backward direction was the resulting number of hydrogen bonds at $R(\text{Co-C}) = 2.0 \text{ \AA}$ equal to 1 or 0.

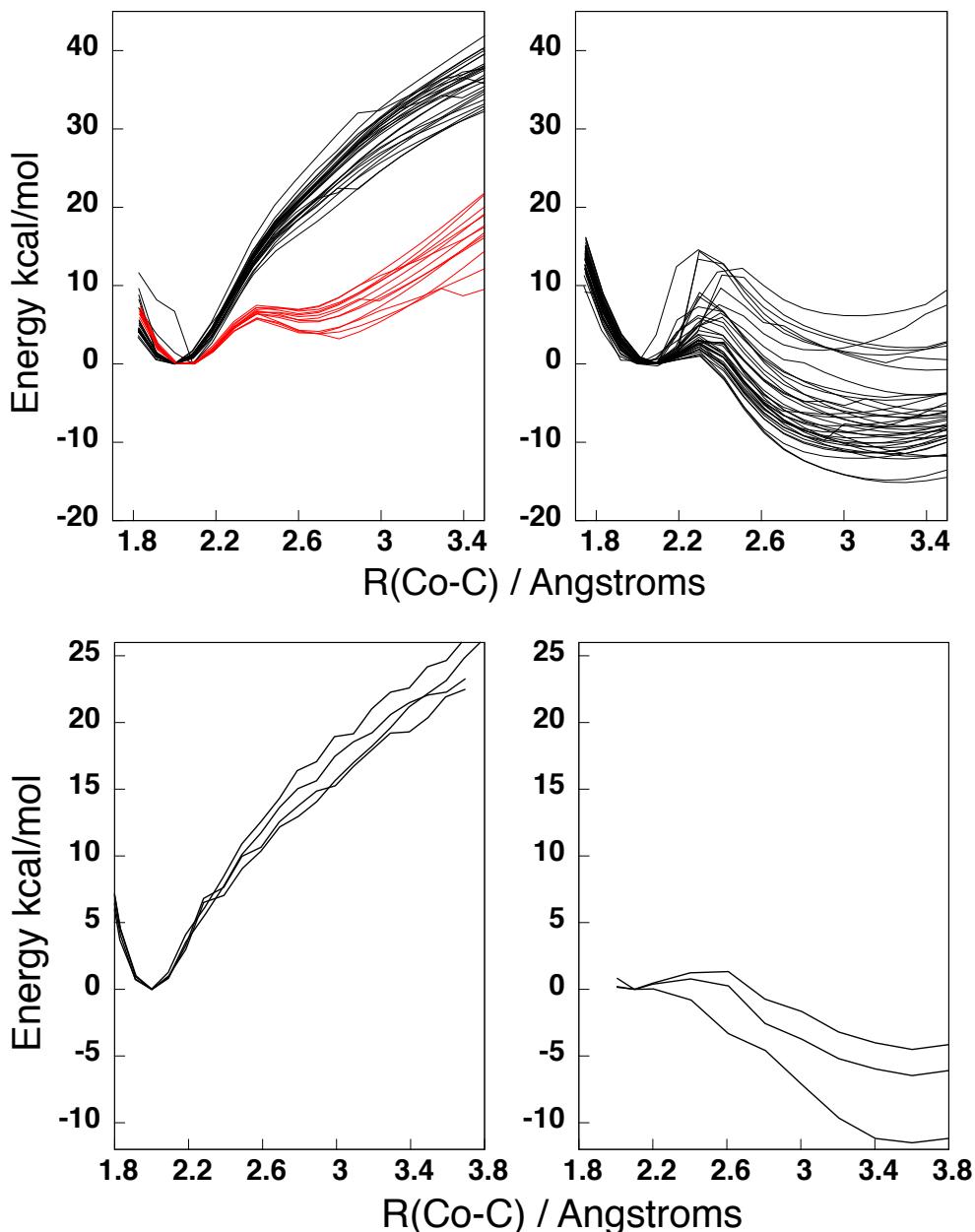


Figure 28 Potential energy profiles for the Co-C bond breaking step in glutamate mutase from adiabatic mapping. The profiles on the top and bottom were generated with EVB and QM/MM potentials respectively (note that they are plotted to different scales). The profiles on the left and right were generated by scanning the $R(\text{Co-C})$ coordinate in the forwards and backwards directions respectively. The profiles plotted in red were all generated starting from structures in which the dihedral angles Φ , θ and χ were far from their equilibrium values for the associated state (see section 4.3.2)

Changes in the dihedral angles Φ , θ and χ during adiabatic mapping were examined in a similar manner to changes in hydrogen bonding. During adiabatic mapping in the forward direction, the values of the dihedral angles never convert between their approximate associated state and dissociated state values. Some structures, (which are unrepresentative of the associated state as discussed earlier) start with dihedral angles close to their dissociated state average values and remain close to them. Most other structures start near their associated state average values and remain close to them. During adiabatic mapping in the backward direction, the value of Co-C4'-C5'-O4' (θ) changes from its dissociated average value to its associated average value in one instance only. This is also the case with the values of N24-Co-C5'-C4' (χ). In no instances does the value of O4'-C1'-N9-C4A (Φ) change from its dissociated average value to its associated average value.

From this analysis we conclude that adiabatic mapping of the R(Co-C) reaction coordinate does not interconvert representative dissociated state and associated state structures and so cannot be accurately modelling the Co-C bond breaking process in enzyme.

The structures from QM/MM adiabatic mapping were not examined quantitatively. However, it was clear from visualising the structures that, as with EVB, adiabatic mapping along the R(Co-C) reaction coordinate with QM/MM did not interconvert representative dissociated state and associated state structures.

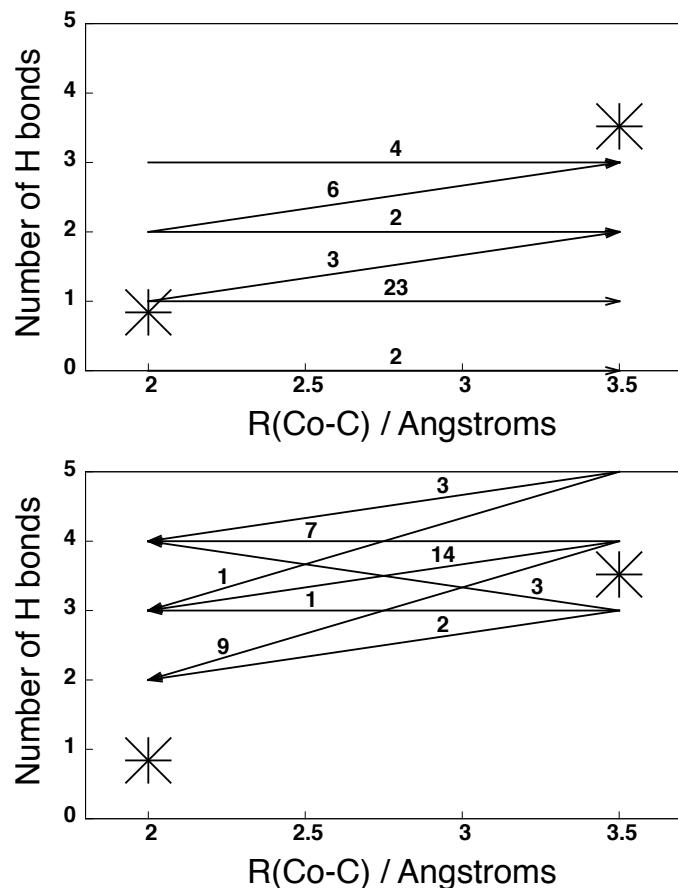


Figure 29 Changes in hydrogen bonding over the course of adiabatic mapping of the $R(\text{Co-C})$ reaction coordinate with EVB. The y axis indicates the number of hydrogen bonds for the structures at $R(\text{Co-C}) = 2 \text{ \AA}$ and 3.5 \AA . The direction of each arrow shows the direction of adiabatic mapping. Above each arrow is the number of profiles that involve the indicated number of hydrogen bonds at the starting and end points. The average numbers of hydrogen bonds from MD simulations of the associated and dissociated states are 0.84 and 3.52 respectively (marked as stars on the plots above).

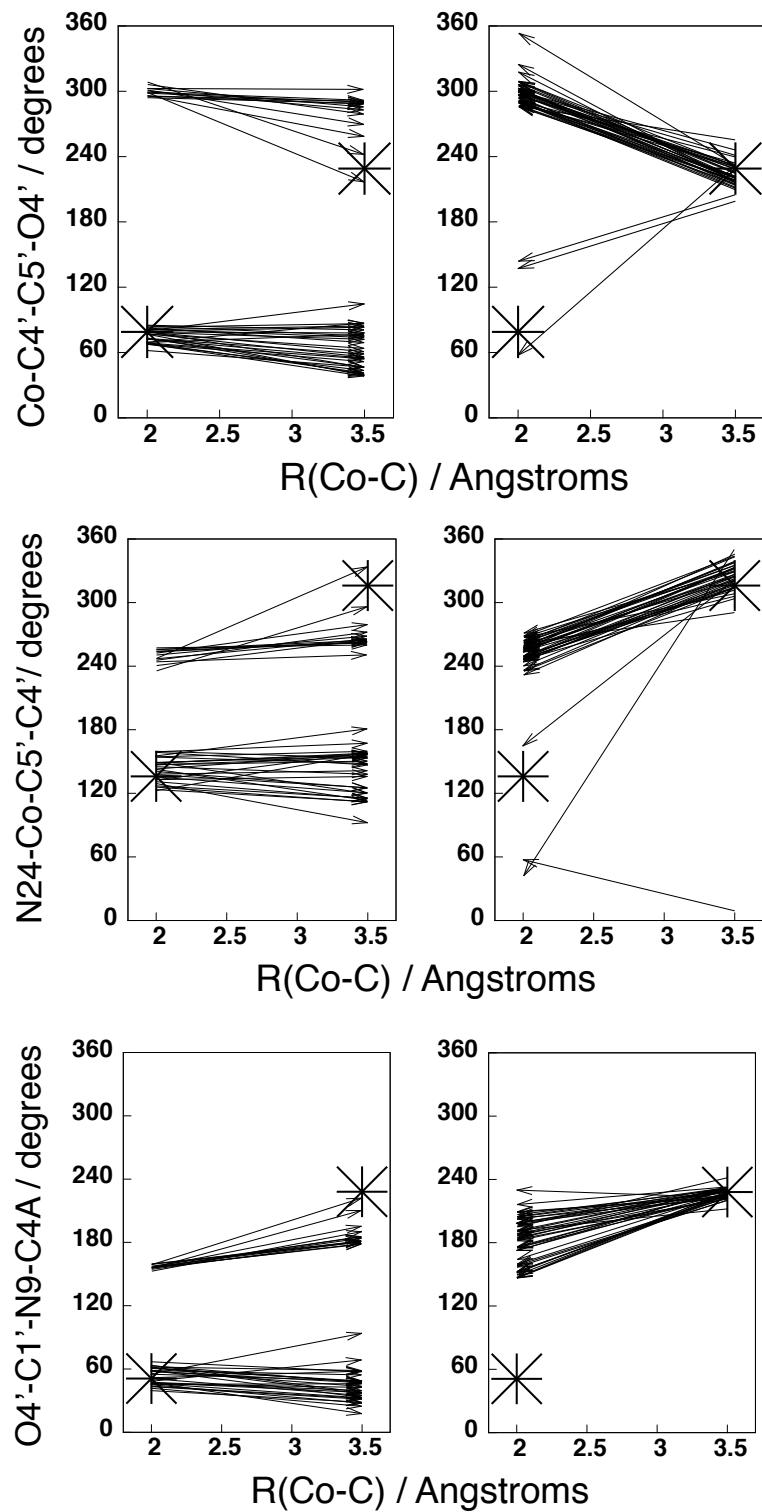


Figure 30 Changes in values of dihedral angles Φ , θ and χ during adiabatic mapping of the $R(\text{Co-C})$ reaction coordinate with EVB. Each arrow represents a single profile. The direction of the arrow indicates the direction of the adiabatic mapping. Stars indicate the average values of each dihedral angle from MD simulations of the associated and dissociated states.

5.3.2. One-dimensional free energy profiles for Co-C bond breaking

Umbrella sampling was carried out along the R(Co-C) reaction coordinate in the forward and backward directions in solution and enzyme. Figure 31 shows the resulting free energy profiles. As with the adiabatic mapping potential energy profiles, the free energy profiles for the Co-C bond breaking in enzyme generated by umbrella sampling in the forward (Co-C bond breaking) direction are qualitatively different from those generated in the backward direction.

The profiles shown here were generated using 400 ps windows. To start with, much shorter windows were used. Free energy profiles were incrementally lengthened in an attempt to increase qualitative convergence with respect to the direction of umbrella sampling. Increasing the length of the windows changed the character of the backward free energy profile, from being exothermic in nature with a very small barrier, (calculated with very short windows) to being endothermic with a barrier of around 10 kcal/mol (generated with long windows as shown in Figure 31). Further increases in length of windows made no significant changes to the backwards-generated free energy profiles.

Increasing the length of windows made no qualitative difference to the free energy profiles calculated with umbrella sampling in the forward direction. These profiles all closely resemble the calculated solution free energy profiles. The solution free energy profiles converged quickly with respect to the length of windows and were insensitive to the direction of umbrella sampling.

As was done with the adiabatic mapping calculations, structural changes over the course of the umbrella sampling simulations were investigated. Average values of the dihedral angles Φ , θ and χ and average number of hydrogen bonds were calculated for the windows $R(Co-C) = 2.0 \text{ \AA}$ and 3.5 \AA . The results are shown in Table 10.

In the case of the simulations in the forward direction, the values for all of the structural parameters investigated at $R(Co-C) = 2.0 \text{ \AA}$ (associated state) were close to their average values taken from the unbiased MD simulations of the associated state. At $R(Co-C) = 3.5 \text{ \AA}$, the values of the dihedrals $D(Co-C4'-C5'-O4')$ and $D(N24-Co-C5'-C4')$ and the average number of hydrogen bonds between adenosyl and Glu300 and between adenosyl and Lys326 were all similar to their average values taken

from the unbiased MD simulations of the dissociated state. However, the hydrogen bonds between the adenosyl base and the protein backbone (Gly68 and Asn123), which were present in the unbiased MD, were not present here. In addition, the average value for the dihedral D(O4'-C1'-N9-C4A) was 35° lower than expected.

In the case of the simulations in the backward direction, the average values of all the structural parameters at $R(\text{Co-C}) = 3.5 \text{ \AA}$ were close to their averages from the unbiased MD simulations of the dissociated state. As the Co-C bond is formed in these simulations the number of hydrogen bonds reduces in agreement with the unbiased MD. The dihedral angles however do not come particularly close to their expected values for the associated state.

As with adiabatic mapping of the $R(\text{Co-C})$ reaction coordinate, umbrella sampling along the same coordinate does not reliably interconvert representative associated and dissociated state structures (although it appears to do slightly better than adiabatic mapping in this regard).

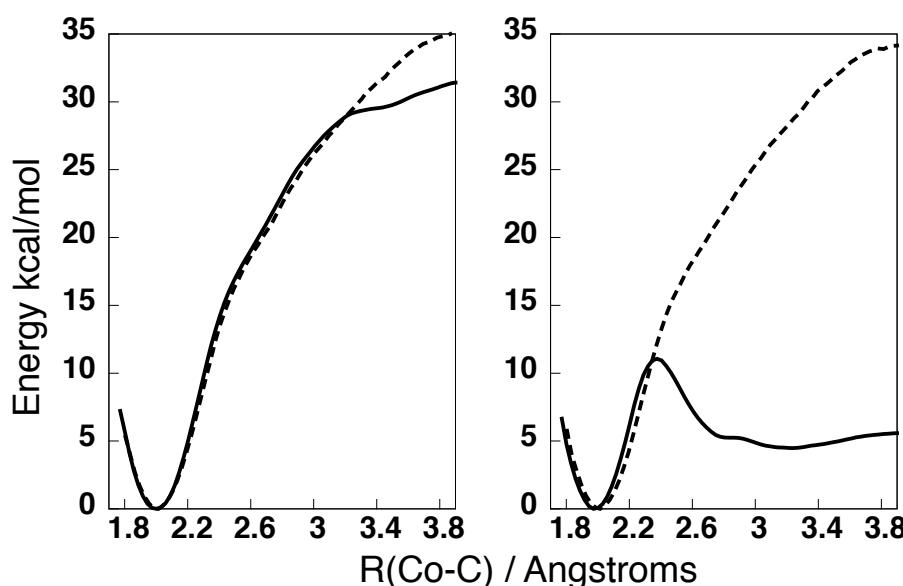


Figure 31 Free energy profiles of the Co-C bond breaking in enzyme (solid lines) and in solution (dashed lines). These were calculated by umbrella sampling along the $R(\text{Co-C})$ reaction coordinate in the forward direction (left) and the backward direction (right).

Table 10 Analysis of changes in active site structure between the associated and dissociated states from umbrella sampling of the R(Co-C) reaction coordinate in the forwards and backwards directions. The structural parameters examined here are those that were used to define the equilibrium associated and dissociated states in section 4.3.2. The values given are average values from the umbrella sampling windows R(Co-C) = 2.0 Å (associated state) and 3.5 Å (dissociated state).

	associated			dissociated		
	forward umbrella	backward umbrella	unbiased MD	forward umbrella	backward umbrella	unbiased MD
330-Ado	1.00	1.00	0.60	1.85	1.93	2.03
326-Ado	0.03	0.28	0.17	0.43	0.33	0.41
base-backbone	0.00	0.00	0.06	0.00	1.08	1.07
total hbond	1.03	1.28	0.84	2.28	3.33	3.52
O4'-C1'-N9-C4A	35	119	54	193	221	228
Co-C4'-C5'-O4'	77	102	79	243	238	229
N24-Co-C5'-C4'	127	77	136	313	310	316

5.3.3. Two-dimensional free energy profiles

As the dihedral angles Φ , θ and χ did not spontaneously convert between their expected associated and dissociated state values with one-dimensional umbrella sampling on the R(Co-C) reaction coordinate, two-dimensional umbrella sampling was tried using these dihedral angles as secondary reaction coordinates.

Figure 32 depicts the two-dimensional free energy surfaces generated by umbrella sampling of the D(Co-C4'-C5'-O4') and R(Co-C) reaction coordinates. Once again, the free energy surface generated by umbrella sampling starting from an associated state structure (from unbiased MD of the associated state) is qualitatively completely different from that starting from a dissociated state structure. The minimum free energy paths taken through the dissociated and associated surfaces closely resemble the corresponding free energy profile generated by one-dimensional umbrella sampling.

As the average values of D(Co-C4'-C5'-O4') from the unbiased MD simulations of the associated and dissociated states were 79° and 229° respectively, the structures were examined at windows D(Co-C4'-C5'-O4') = 75° , R(Co-C) = 2.0 Å and D(Co-C4'-C5'-O4') = 235° , R(Co-C) = 3.6 Å. A summary of the structural changes is given in Table 11. This two-dimensional umbrella sampling procedure appears to interconvert the associated and dissociated state structures with the notable exception of

the dihedral angle D(O4'-C1'-N9-C4A), which does not arrive near its expected dissociated state value when umbrella sampling in the forward direction or at its expected associated state value when simulating in the backward direction.

Figure 33 depicts the two-dimensional free energy surfaces generated by umbrella sampling of the D(O4'-C1'-N9-C4A) and R(Co-C) reaction coordinates. As the average values of D(O4'-C1'-N9-C4A) from the unbiased MD simulations of the associated and dissociated states were 51° and 228° respectively, the structures were examined at windows D(O4'-C1'-N9-C4A) = 50°, R(Co-C) = 2.0 Å and D(O4'-C1'-N9-C4A) = 230°, R(Co-C) = 3.5 Å. A summary of the structural changes is given in Table 12. Again, the surfaces generated by umbrella sampling in the forward direction do not match those generated in backwards direction. Umbrella sampling along these two coordinates was not able to convert a dissociated state starting structure into a reasonable associated state structure. The associated state structure was however converted into a reasonable dissociated state structure (at least according to all of the structural parameters shown in Table 12).

Given this successful structural conversion and the plausible minimum energy path for Co-C dissociation taken through the free energy surface of the backwards simulation, it is tempting to think that a reasonable model of Co-C bond breaking in enzyme has been found. However, the resultant free energy surface is inconsistent with the unbiased MD simulations of the associated state. The lowest energy value of the dihedral angle D(O4'-C1'-N9-C4A) in the associated state region of this surface is around 200°, whereas the unbiased MD indicates that the equilibrium value should be around 75°.

To further investigate this inconsistency, one-dimensional umbrella sampling using the dihedral angle D(O4'-C1'-N9-C4A) was performed using the associated state (non-EVB) force field. Figure 34 (left) shows free energy profiles from one-dimensional umbrella sampling of this dihedral angle (black lines), starting from different structures taken from unbiased MD simulations of the associated state. Also shown are one-dimensional slices from the two-dimensional free energy surfaces from simulations in the forward direction (red) and the backward direction (blue) at R(Co-C) = 2.0 Å. While the profiles from one-dimensional umbrella sampling are not converged with respect to the starting structure used, they all have a minimum value between 50° and 75° consistent with the average value from the unbiased MD. The one-dimensional slice from the two-dimensional umbrella sampling simulations in the forward direction is broadly

consistent with these profiles whereas the one-dimensional slice from the backward simulation has a completely different character.

Similar simulations were performed using the dissociated state force field and starting from structures taken from the unbiased MD simulations of the dissociated state. The profiles are shown in Figure 34 (right). A similar picture emerges, where the slice taken at $R(\text{Co-C}) = 3.5 \text{ \AA}$ from the backward surface broadly agrees with the profiles from one-dimensional umbrella sampling, whereas the slice taken from the forward surface is entirely different.

It seems that in all of the simulations of the Co-C bond breaking presented here there is a systematic bias towards the state in which the simulation was started. Analysis of the unbiased MD simulations in section 4.3 showed that the active site undergoes significant structural rearrangement when converting between the associated and dissociated states and that this rearrangement occurs on a timescale of nanoseconds. As this will involve the crossing of numerous energy barriers in degrees of freedom orthogonal to the reaction coordinate (e.g. the dihedral angles already discussed) it is not surprising that adiabatic mapping is unable to model this reaction. This is because, as a minimisation technique, adiabatic mapping generally won't cross barriers orthogonal to the reaction coordinate.

Umbrella sampling does not suffer from this problem as (given sufficient sampling) barriers in degrees of freedom orthogonal to the reaction coordinate can be crossed (provided their heights are within the range of thermally accessible energies). As discussed in section 4.3, the crystal structure is thought to resemble the dissociated state more than the associated state. This is consistent with the low RMSDs of the backbone atoms for the unbiased single state MD simulations of the dissociated state and the high RMSDs for the equivalent associated state simulations. This difference in RMSDs shows that there are significant structural differences between the associated and dissociated states beyond just the conformation of the adenosyl moiety and its interactions with the active site residues. It seems plausible that the reason for the systematic bias towards the state in which the umbrella sampling simulations were started could be related to the stochastic boundary conditions used. The restraints imposed on the protein structure may not allow sufficiently facile rearrangement of the active site structure necessary to interconvert between associated and dissociated states. In order to investigate this possibility further and possibly address the issue, these umbrella sampling simulations would ideally be repeated with the full system, no positional restraints and periodic

boundary conditions. This is currently not possible because the system size for these simulations would be greater than the current maximum system size allowable in the current implementation of EVB in CHARMM. The factor limiting the maximum system size is the memory required to store the arrays of the combined forces of each EVB state (which are of size ‘number of atoms x number of EVB states’). This could in principle be addressed with some fairly trivial restructuring of the EVB code in CHARMM but goes beyond the scope of this work.

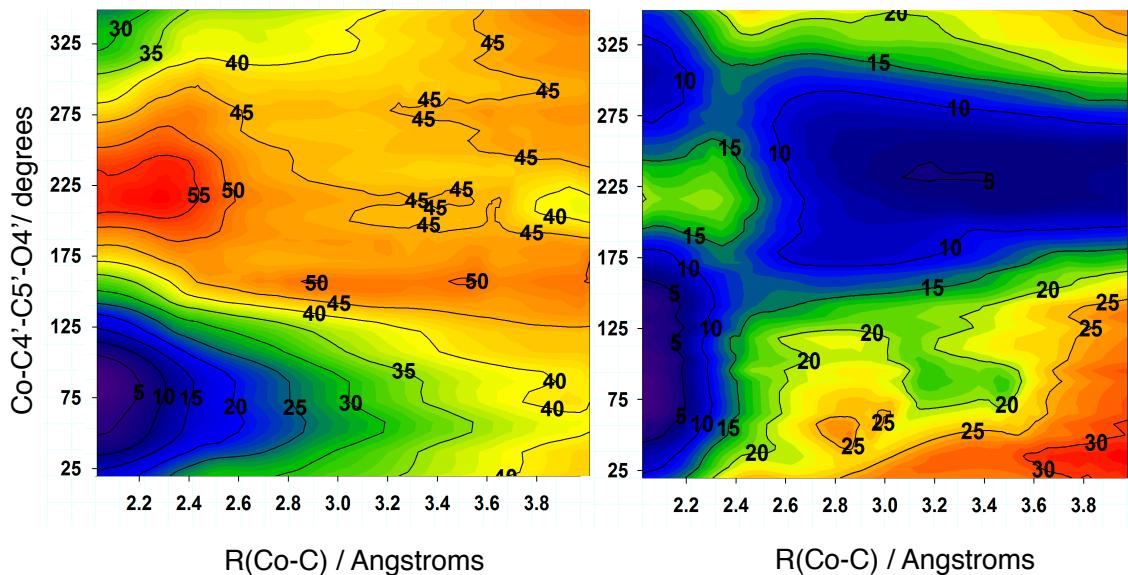


Figure 32 Free energy surfaces for Co-C bond breaking in enzymes from 2D umbrella sampling using $R(\text{Co-C})$ and $D(\text{Co-C}4'\text{-C}5'\text{-O}4')$ as reaction coordinates. Simulations starting in the associated and dissociated states generated the surfaces on the left and right respectively.

Table 11 Analysis of changes in active site structure between the associated and dissociated states from 2D umbrella sampling of the R(Co-C) and D(Co-C4'-C5'-O4') reaction coordinates. The values given are average values from the umbrella sampling windows R(Co-C) = 2.0 Å, D(Co-C4'-C5'-O4') = 75° (associated state) and R(Co-C) = 3.6 Å, D(Co-C4'-C5'-O4') = 235° (dissociated state).

	associated			dissociated		
	forward umbrella	backward umbrella	unbiased MD	forward umbrella	backward umbrella	unbiased MD
330-Ado	1.00	0.00	0.60	1.92	2.00	2.03
326-Ado	0.00	0.62	0.17	0.31	0.54	0.41
base-backbone	0.00	0.08	0.06	0.85	0.77	1.07
total hbond	1.00	0.69	0.84	3.08	3.31	3.52
O4'-C1'-N9-C4A	36	127	54	130	219	228
Co-C4'-C5'-O4'	77	76	79	235	236	229
N24-Co-C5'-C4'	131	141	136	306	311	316

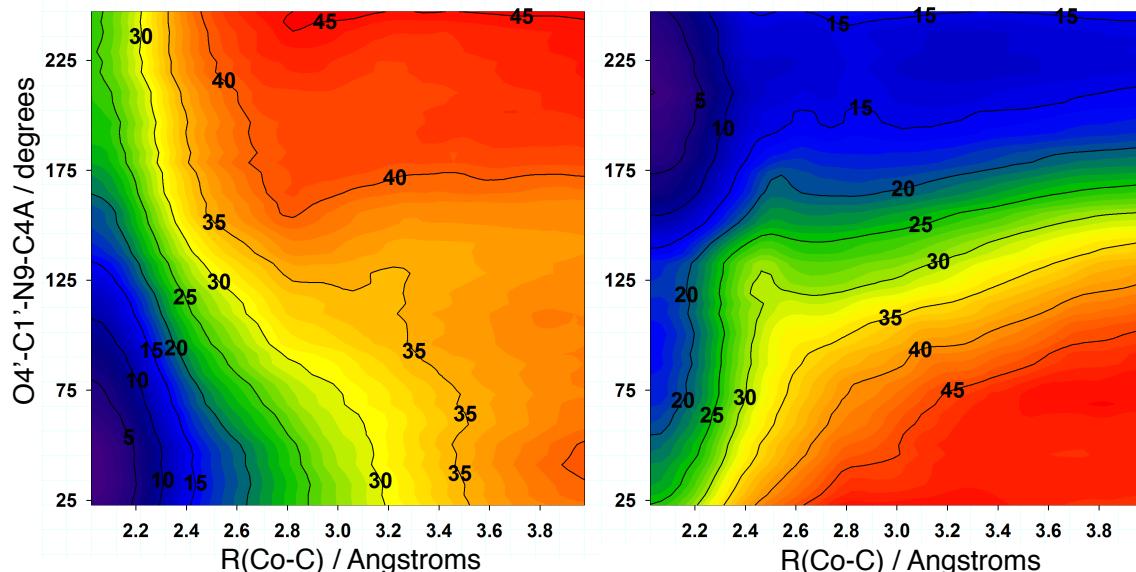


Figure 33 Free energy surfaces for Co-C bond breaking in enzymes from 2D umbrella sampling using R(Co-C) and D(O4'-C1'-N9-C4A) as reaction coordinates. Simulations starting in the associated and dissociated states generated the surfaces on the left and right respectively.

Table 12 Analysis of changes in active site structure between the associated and dissociated states from 2D umbrella sampling of the R(Co-C) and D(O4'-C1'-N9-C4A) reaction coordinates. The values given are average values from the umbrella sampling windows R(Co-C) = 2.0 Å, D(O4'-C1'-N9-C4A) = 50° (associated state) and R(Co-C) = 3.5 Å, D(O4'-C1'-N9-C4A) = 230° (dissociated state).

	associated			dissociated		
	forward umbrella	backward umbrella	unbiased MD	forward umbrella	backward umbrella	unbiased MD
330-Ado	1	0.00	0.60	2.06	1.72	2.03
326-Ado	0	0.00	0.17	0.17	0.44	0.41
base-backbone	0	0.00	0.06	1.33	1.00	1.07
total hbond	1	0.00	0.84	3.56	3.17	3.52
O4'-C1'-N9-C4A	48	54	54	230	229	228
Co-C4'-C5'-O4'	76	85	79	226	231	229
N24-Co-C5'-C4'	132	204	136	320	318	316

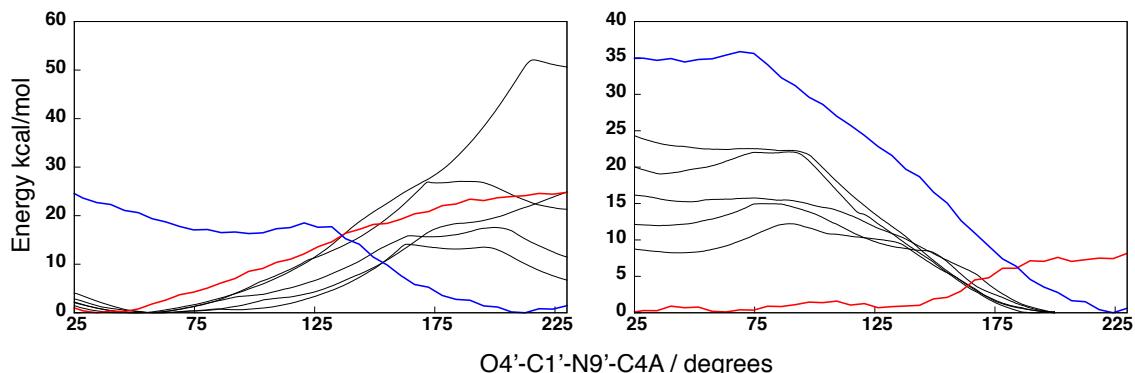


Figure 34 Free energy profiles for the rotation about the dihedral angle D(O4'-C1'-N9-C4A) in the associated (left) and dissociated (right) states. Profiles in black are from 1D umbrella sampling with the appropriate single state (non-EVB) force field. Profiles in red and blue are slices from the relevant 2D free energy surfaces shown in Figure 33 at R(Co-C) = 2.0 Å and 3.5 Å respectively.

5.3.4. Modelling hydrogen abstraction

Adiabatic mapping was used to study the hydrogen abstraction step with both QM/MM and EVB potentials for the wild type system and all of the mutants studied in section in Chapter 4. Figure 35 shows a number of potential energy profiles generated by adiabatic mapping with both methods (for the wild type). Table 13 summarises the statistical properties of the pathways calculated for each system and compares them to the experimental rate constants for tritium exchange from the experiments of our collaborators in the Marsh group.¹¹⁸

The profiles calculated with EVB have similar properties to those calculated with QM/MM except that the QM/MM barriers are generally lower than the EVB barriers. This is to be expected as our calculations in section 3.1, which showed that DFT underestimates the barrier to this hydrogen transfer (compared to CCSD(T)-F12).

As discussed in section 1.2.3, the barrier heights obtained by adiabatic mapping starting from different structures are often very sensitive to the starting structure used. The range and standard deviation of barrier heights are fairly low in the case of the wild type for this step, but they are much higher for some of the mutants. The exponential averaging method was used to generate a best estimate for the barrier to reaction from all pathways for each system (see Table 13). The standard deviations of the barrier heights within each system are fairly high relative to the differences in exponentially averaged barrier heights between the systems. Because of this we cannot be confident that we have converged with respect to the exponentially averaged barrier heights for each system relative to the others. It would not be surprising for example if adding a further 50 pathways to each system would significantly change the ranking of these systems in terms of exponentially averaged barrier heights for this hydrogen abstraction step.

Figure 36 plots the exponentially averaged barrier heights against the log of the relative rate of tritium exchange. It is important to note that we would not expect even a set of hypothetically perfect calculated barrier heights for the hydrogen abstraction step to correlate perfectly with the log of the rate of tritium exchange. The Co-C bond breaking and hydrogen abstraction steps are known to be kinetically coupled.⁷⁴ As we expect these mutations to also affect the rate of Co-C bond breaking (by unknown amounts) we can only hope to partially rationalise the observed differences in rate of tritium exchange by studying the hydrogen abstraction step alone. Ideally we would have calculated barrier heights for Co-C bond breaking in all systems, however given our failure to do this convincingly for the wild type system, we have not attempted it for the mutant systems. With this and our reservations about convergence in mind we can tentatively observe that there does appear to be some relationship between the exponentially averaged barrier heights and the rates of tritium exchange. One notable anomaly is E330D, which has a similar calculated barrier height to the wild type despite its much slower rate of tritium exchange and longer average C5'-HG2 distance in MD simulations. This is not consistent with our hypothesis (outlined in section 4.3.5) that those mutants with larger average C5'-HG2 distances had higher barriers for the hydrogen abstraction step, partially explaining their relative tritium exchange rates.

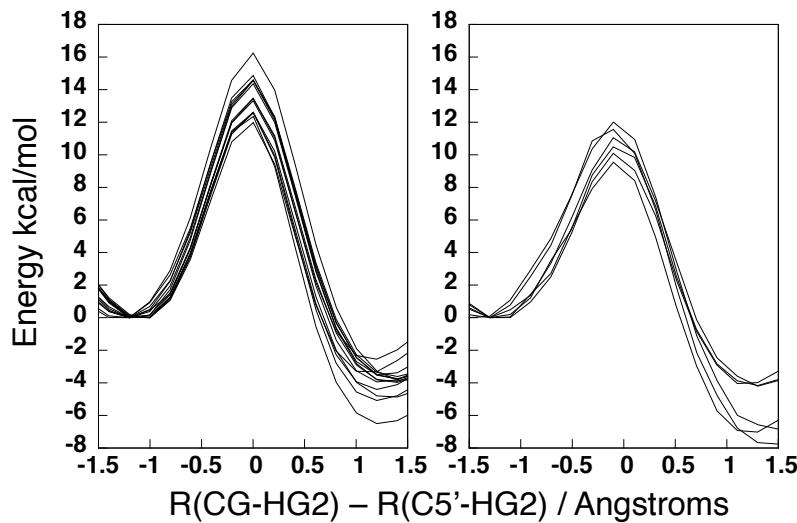


Figure 35 Potential energy profiles for the hydrogen abstraction step in wild type glutamate mutase from adiabatic mapping with EVB (left) and QM/MM (right). Each profile was generated starting with a different structure taken from MD simulations of the dissociated state.

Table 13 Statistical properties of the set of potential energy profiles generated by adiabatic mapping for each system. Exp. av. is the exponential average of the calculated barrier heights. $\log(k_T)$ is the log of the experimental tritium exchange rate constant.¹¹⁸ All values are in units are kcal/mol (except the number of paths and $\log(k_T)$).

	paths	min	max	mean	SD	exp. av	$\log(k_T)$
wild QMMM	6	9.5	12.0	10.8	0.9	10.3	0.0
wild EVB	14	12.0	16.2	13.7	1.2	12.9	0.0
E330D EVB	12	11.7	22.5	19.2	2.9	13.1	-2.4
E330Q EVB	16	16.2	21.8	18.3	1.6	17.2	-2.0
E330A EVB	16	16.6	22.9	19.8	1.9	18.0	-3.7
K326Q EVB	13	17.7	25.7	21.5	2.3	19.1	-3.9
K326M EVB	11	16.6	24.0	19.6	2.7	17.4	-2.8

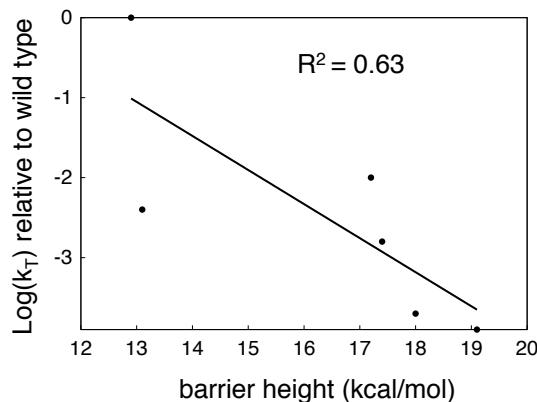


Figure 36 The exponentially averaged barrier heights (from adiabatic mapping) plotted against the log of the experimental tritium exchange rate constants for each system.

5.3.5. Umbrella sampling of hydrogen abstraction

As already discussed, it can be difficult to obtain reliable and converged barrier heights from adiabatic mapping. In an attempt to improve on the rather approximate potential energy barrier heights calculated in the previous section, umbrella sampling simulations were performed to generate free energy profiles of the hydrogen abstraction step.

Figure 37 shows a number of free energy profiles for the hydrogen abstraction step calculated with umbrella sampling, each starting from different structures. Table 14 shows the statistical properties of the set of calculated barrier heights for each system. Also given is the free energy for each system calculated by performing WHAM on the combined data from all simulations for that system. This will be used as our best estimate of the free energy for the hydrogen abstraction step and will be referred to as the calculated free energy barrier for each system. The range and standard deviations of the free energy profiles for each system are smaller than for their equivalent potential energy profiles. While still not tightly converged, the calculated free energy barriers are certainly more convincingly converged than the exponentially averaged barrier heights from adiabatic mapping.

Having plotted the free energy barriers against the log of the tritium exchange rates (Figure 38), we observe a more convincing correlation than with the potential energy barriers. A plausibly significant relationship can also be observed when the calculated free energy barriers are plotted against the average C5'-HG2 distances from MD simulations. The points on both of these plots can be divided into three groups. Group 1 contains only the wild type, which has the lowest average C5'-H distance and by far the lowest barrier to hydrogen abstraction and rate of tritium exchange. Group 3 contains the E330A and K326Q mutants. These mutants have the largest calculated free energy barrier heights, the longest average C5'-H distances and the slowest rates of tritium exchange. Group 2 contains the other mutant systems E330D, E330Q and K326M. Group 2 has intermediate values of these three properties, which are fairly closely clustered, with the exception of the anomalously low average C5'-H distance of the E330Q system.

Umbrella sampling simulations of the adenosyl radical and glutamate in a box of TIP3P molecules were performed to calculate the free energy barrier to hydrogen transfer

in solution. For technical reasons of still unknown origin, these simulations tended to crash when umbrella sampling windows of longer than 10 ps were used. The barrier height obtained with 5 ps and 10 ps windows were 16.3 kcal/mol and 17.5 kcal/mol respectively. Considering that these values are not converged and that there is clearly some technical problem with the simulations we should be careful in our interpretation of these numbers. Having said this, it is worth noting that these calculated barrier heights are of similar magnitude to the enzymatic barrier heights calculated with umbrella sampling. This suggests that the enzyme may not actually lower the barrier to this step at all, rather the catalytic effect may be due to the reduction of barriers to other steps.

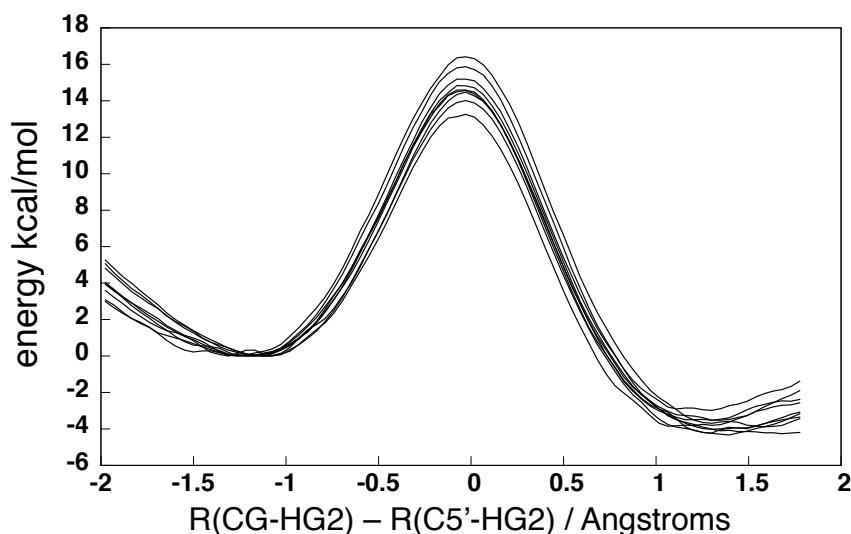


Figure 37 Free energy profiles of the hydrogen abstraction step in glutamate mutase generated from umbrella sampling simulations. Each profile was generated started from a different structure taken from MD simulations of the dissociated state.

Table 14 Statistical properties of the set of free energy profiles generated by umbrella sampling simulations of each system. The values in the ‘combined’ column are the free energy barriers calculated using the data from all of the umbrella sampling simulations of that system. Average distances are in Å. All other units are in kcal/mol (except $\ln(k_T)$).

	min	max	mean	SD	comb.	$\ln(k_T)$	av dist
wild	13.3	16.4	14.8	0.9	14.7	0.0	2.67
E330D	17.9	20.5	19.0	0.9	19.0	-2.4	3.27
E330Q	17.9	21.3	19.4	1.0	19.3	-2.0	2.85
E330A	19.0	24.3	21.1	1.4	21.0	-3.7	3.61
K326Q	18.9	21.9	20.2	1.0	20.2	-3.9	3.43
K326M	15.8	19.9	18.8	1.2	18.7	-2.8	3.30

Having estimated the barriers to hydrogen abstraction for several systems with both exponentially averaged potential energy barriers from adiabatic mapping and free energy barriers from umbrella sampling, we can check for consistency between these methods. Because the same set of starting structures and same potential has been used, this should be a reasonably fair comparison. Figure 40 plots the barrier heights calculated with one method against the other. The resulting correlation between the two methods is not strong and disappears completely if the wild type data point is removed from the set. We believe that the results of the umbrella sampling simulations are more accurate than those from adiabatic mapping. This is not only because the results are better converged, but because umbrella sampling has numerous theoretical advantages over adiabatic mapping such as a much better exploration of configurational space orthogonal to the reaction coordinate and inclusion of entropic effects. Given the assumed superiority of umbrella sampling and the poor correlation between the two methods, we can conclude that adiabatic mapping is not able to give barrier heights that are accurate enough for meaningful comparison between the systems studied here.

Because we do not have any experimental rates for the hydrogen abstraction elementary step, we cannot confidently comment on the accuracy of the barriers calculated by umbrella sampling. The barriers produced are however broadly consistent with our interpretation of the experimental rates of tritium exchange and the average C5'-H distances from MD simulations.

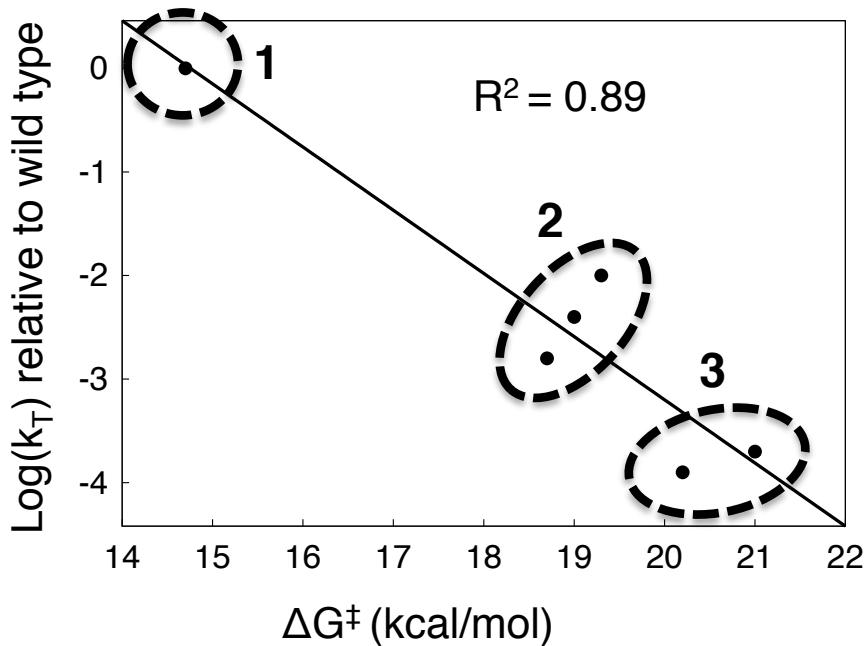


Figure 38 Free energy barrier heights for the hydrogen abstraction step calculated with umbrella sampling plotted against the log of the experimental tritium exchange rate constants for all systems.

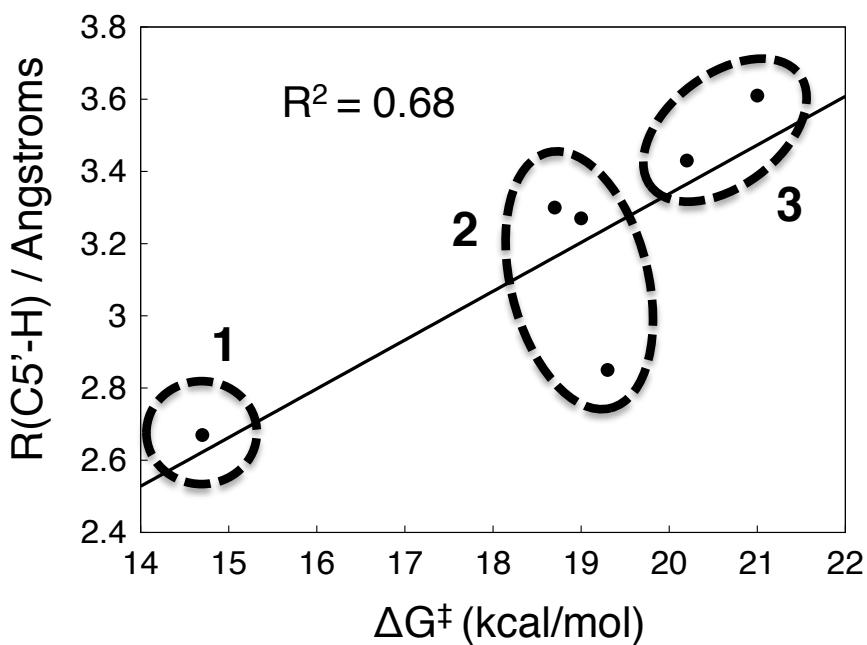


Figure 39 Free energy barrier heights for the hydrogen abstraction step calculated with umbrella sampling plotted against the average C5'-H distances from MD simulations for all systems.

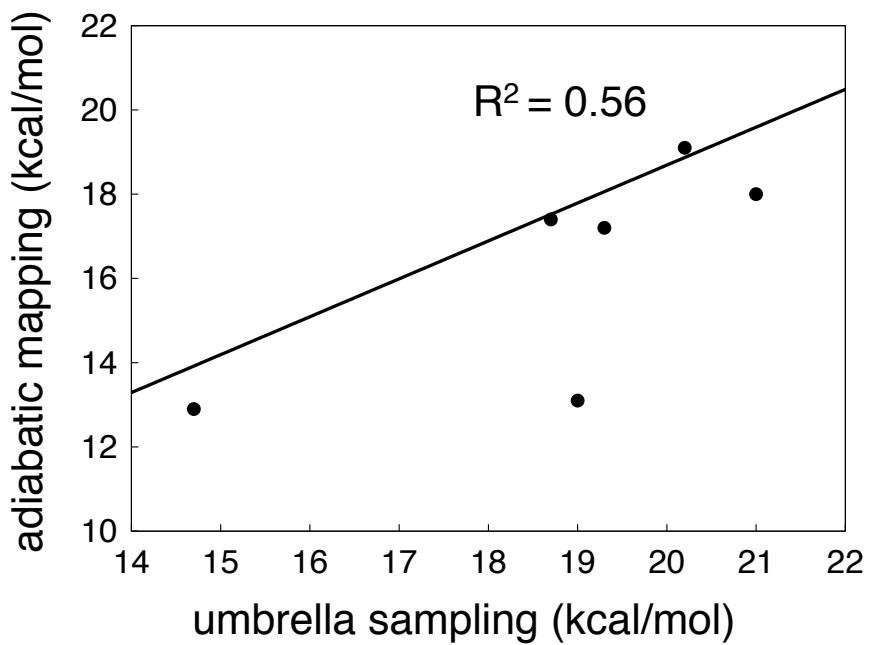


Figure 40 Free energy barrier heights for the hydrogen abstraction step for each system calculated with umbrella sampling plotted against the exponentially averaged potential energy barrier heights from adiabatic mapping.

Chapter 6. Glutamate mutase conclusions

6.1. Conclusions

Adenosylcobalamin dependent enzymes are remarkable in their ability to generate and control carefully highly reactive organic radical species in order to catalyse difficult carbon skeleton rearrangement reactions. The first step in the reactions of all enzymes in this class is the homolytic cleavage of the Co-C bond to generate a 5'-deoxyadenosyl radical. This species initiates further radical chemistry by abstracting a hydrogen from a substrate or protein residue in a highly specific manner. How these enzymes accelerate Co-C bond breaking by an estimated 12 orders of magnitude relative to solution⁶¹, whether the adenosyl radical exists as a metastable or transient intermediate and how this highly reactive species is directed to react so specifically are the fundamental questions that have driven the work described here. As one of the most rigorously experimentally studied adenosylcobalamin dependent enzymes, glutamate mutase was chosen as the subject of the present molecular modelling study with the hope of providing further insight into the chemistry of the class as a whole.

A growing body of literature has emerged indicating that Co-C bond breaking is accompanied by significant changes in the conformation of the adenosyl moiety, the active site and possibly even larger sections of protein.^{33,85,93,117} It has also been proposed that the adenosyl moiety increases its hydrogen bonding with the protein as the Co-C bond is broken and that the resulting electrostatic stabilisation accounts for the catalytic effect of these enzymes.^{37,117} In order to characterise these conformational changes and hydrogen bonding patterns, periodic boundary condition molecular dynamics simulations of the full solvated holoenzyme with substrate bound were performed in both the associated (Co-C formed) and dissociated (Co-C broken) states.

The crystal structure of glutamate mutase used in this work contains a superposition of two conformations for the adenosyl moiety (A and B), both of which clearly have broken Co-C bonds. Average structures from the MD simulations of the dissociated state exhibit adenosyl and active site conformations that are extremely similar to crystal

structure conformation A. The RMSD of the protein backbone positions relative to the crystal structure were low and stable throughout the simulations of the dissociated state. Taken together, this indicates that conformation A is representative of the dissociated state and that the force field used performs well in modelling the structural properties of this state. In the dissociated state, the adenosyl radical forms a number of stable hydrogen bonds with active site amino acid residues. The $-NH_2$ group of the adenosyl adenine base hydrogen bonds with the backbone carbonyl oxygens of Gly68 and Asn123. The two adenosyl ribose $-OH$ groups both hydrogen bond with Glu330. Lys326 also hydrogen bonds to the ribose $-OH$ groups. In addition to the hydrogen bonding patterns, the conformation and orientation of the adenosyl moiety was defined by three dihedral angles (Φ , θ and χ) and the ribose pseudorotational conformation. These dihedral angles were restricted to a fairly limited range of values throughout the simulations, which are close to their respective values in crystal structure conformation A. The pseudorotational conformation was almost entirely restricted to C3' endo, which is also the conformation found in crystal structure conformation A. The distance between the C5' radical centre and the abstractable hydrogen of the substrate was monitored throughout the simulations and found to be restricted to a fairly narrow range of small values. In the dissociated state, the network of robust hydrogen bonds restricts both the internal conformation of the adenosyl group and its relative orientation in the active site, with the result that the C5' radical centre is directed towards and brought into close contact with the hydrogen to be abstracted from the substrate.

As the adenosyl conformation in the dissociated state is well represented by crystal structure conformation A, the analysis contained in the above paragraph arguably contributes little insight that could not have been gained by study of the crystal structure alone. In contrast, prior to performing these MD simulations, the active site structure and adenosyl conformation in the associated state was unknown. The structures of both active sites from both simulations of the associated state converged with respect to the hydrogen bonding patterns and dihedral angles Φ , θ and χ . The average number of adenosyl-protein hydrogen bonds was reduced from 3.5 in the dissociated state to 0.8 in the associated state. This significant change in number of hydrogen bonds is consistent with the hypothesis that electrostatic stabilisation of the dissociated state is a major catalytic effect in these enzymes.⁶⁴ The values for the dihedral angles were spread over fairly narrow ranges that were very different from the dissociated state ranges. A much wider range of pseudorotational conformations was accessible than in the dissociated state. Whilst the

hydrogen bonding converged quickly, it took up to 13 ns of simulation for the adenosyl to relax to its equilibrium internal conformation. In the start of all simulations of the associated state, it appears that the system was trapped in a metastable conformation as an artefact of starting from a crystal structure that was not representative of the equilibrium associated state. The RMSDs of the protein backbone positions relative to the crystal structure were much higher than in the dissociated state and had not levelled off by the end of the simulations. This suggests that there may be larger scale structural differences between the two states and that the simulations of the associated state may not have completely relaxed into an equilibrium ensemble of structures.

On the basis of the analysis presented here, it can be confidently said that significant conformational changes accompany Co-C bond breaking in terms of the adenosyl's orientation, internal conformation and interactions with the protein. There is evidence that this may be linked to larger scale changes in the protein conformation. It is proposed that the observed conformational changes are driven by an active site preorganised to tightly bind the dissociated adenosyl radical in a specific conformation. This appears to serve the dual purpose of electrostatically stabilising the dissociated state and of carefully directing the C5' radical centre to abstract a specific hydrogen from the substrate.

The behaviour of the mutant systems was largely consistent with that of the wild type in terms of the changes in conformation and hydrogen bonding associated with Co-C bond breaking. While the difference in the number of hydrogen bonds between the associated and dissociated states varied between the mutants, the estimated variation in electrostatic stabilisation of the dissociated state that this implies was not sufficient to rationalise their relative experimental rates of tritium exchange. However, a correlation was found between the average distance between the C5' radical centre and the abstractable hydrogen of the substrate. This led to the hypothesis that a reduced ability to optimally orientate the adenosyl radical for hydrogen abstraction may be a contributing factor to the reduction in activity of the mutant systems.

Perhaps significantly, the dissociated adenosyl radical in substrate free simulations does not form as many hydrogen bonds with the protein as in the wild type. This suggests that the active site is less able to stabilise the dissociated state in the absence of substrate, with the result that the adenosyl radical is not produced unless required for turnover. This is consistent with the hypothesis that substrate binding triggers some sort of conformational change in the active site that facilitates Co-C bond breaking.¹²⁰

The Co-C bond dissociation energy of a minimal gas phase model of adenosylcobalamin was calculated with high level *ab initio* calculations. Out of a range of DFT methods tested, B3LYP* with Grimme's empirical dispersion correction version 2 (B3LYP*-D2) was found to best reproduce the *ab initio* results. Consequently this was deemed to be the most appropriate DFT method for modelling adenosylcobalamin chemistry (in agreement with some^{19,80}, but not all recently published work addressing the same issue⁸¹).

Given the conclusion from MD simulations that significant conformational changes accompany Co-C bond breaking in enzyme, it has become clear that sufficient conformational sampling is essential for accurate modelling of this process. In order to attempt this, analytic potential energy surfaces for the Co-C bond breaking and hydrogen abstraction steps were developed using the empirical valence bond method (EVB). These potentials were carefully calibrated to gas phase DFT and *ab initio* reference data and were found to produce potential energy profiles for both steps that are consistent with those calculated using QM/MM (DFT). Importantly, these potentials are computationally cheap enough to permit several nanoseconds of sampling of an enzyme system on a timescale of days.

Potential energy profiles for Co-C bond breaking in glutamate mutase were calculated by adiabatic mapping along the Co-C distance reaction coordinate on the EVB and QM/MM potential energy surfaces. This was done in the forward (bond breaking) and backward (bond forming) directions starting from a range of structures taken from MD simulations of the associated and dissociated states respectively. The profiles obtained from adiabatic mapping in the forward direction closely resembled the equivalent gas phase profiles, exhibiting no significant reduction in the energy required to break the Co-C bond. In contrast, the characteristics of the profiles obtained in the backwards direction varied from having almost no barrier and an exothermic reaction energy for Co-C bond breaking to having a barrier of around 12 kcal/mol and a reaction energy of around +5 kcal/mol. These two sets of profiles are completely inconsistent with each other. Analysis of the hydrogen bonding and adenosyl conformation (via the dihedral angles Φ , θ and χ) along these profiles showed that adiabatic mapping in either direction did not interconvert representative associated and dissociated state structures. Given the significant conformational changes orthogonal to the Co-C distance reaction coordinate that must occur during Co-C bond breaking, it is proposed that this process cannot be properly modelled with adiabatic mapping along this coordinate alone and

possibly even with any combination of reaction coordinates. The results of all previous studies that have employed this method should be sceptically interpreted with this in mind.^{63,84,122}

It was hoped that umbrella sampling, which (unlike adiabatic mapping) is able to cross barriers in degrees of freedom orthogonal to the reaction coordinate, would allow for the necessary conformational changes to occur during Co-C bond breaking. Unfortunately, the free energy profiles calculated from one-dimensional umbrella sampling along the Co-C distance in the forward and backward directions were qualitatively very different from each other (as was the case with the potential energy profiles from adiabatic mapping). Again, structural analysis showed that representative associated and dissociated state structures were not being properly interconverted. With two-dimensional umbrella sampling of the R(Co-C) and Φ reaction coordinates, starting from the associated state, it did prove possible to force the adenosyl radical into the correct dissociated state conformation with respect to the hydrogen bonding and dihedral angles Φ , θ and χ . However, the resulting free energy surface was inconsistent with the MD simulations of the associated state in that the minimum for Φ at short Co-C distances was much closer to the dissociated state than the associated state region. It seems that there is some aspect of the simulation protocol used here that is significantly biasing the system towards its initial state.

The unbiased periodic boundary MD simulations of the whole protein system in the associated state showed that several nanoseconds of sampling at least is required for the system to relax from its starting structure (resembling the dissociated state) to an equilibrium ensemble of structures. One hypothesis as to why the system is biased towards its starting structure in the umbrella sampling simulations is that the active site structure may not be able to respond properly to changes brought about by the breaking of the Co-C bond. It could be that even the extensive sampling performed here was insufficient. However, as the results did not seem to change significantly with extended sampling, it seems more likely that the restraints on the protein structure imposed by the stochastic boundary conditions are too restrictive.

Three studies of methylmalonyl-CoA mutase have used sampling techniques to calculate free energy profiles for Co-C bond breaking and hydrogen abstraction.^{33,64,93} One of these (the EVB study) modelled a spherical subsystem of radius 18 Å with positional restraints on the boundary and so may also suffer from over restrictive conditions. The two Car-Parinello studies were performed on the full system with

periodic boundary conditions, but involved comparatively limited and possibly insufficient sampling (30 – 45 ps for Co-C breaking).^{33,93} One of the studies³³ provides a movie of Co-C bond breaking in the supporting information, in which it can be seen that some conformational changes do occur, such as the change in ribose pseudorotation and the increase in the number of hydrogen bonds between the adenosyl –OH groups and the active site glutamate from 1 to 2. However, the adenosyl adenine base barely moves during the reaction and the dihedral angle θ appears to be in a highly strained conformation in the associated state, which is inconsistent with the MD results presented here.

Calculations of the potential and free energy barriers to hydrogen transfer were far more straightforward and appeared to have converged to within a few kcal/mol. The best estimate for the free energy barrier and reaction energy as calculated from umbrella sampling were 14.7 kcal/mol and -4.0 kcal/mol respectively. The barriers for hydrogen transfer calculated in enzyme, solution and gas phase were all similar, suggesting that the enzyme does not significantly reduce the barrier to this step as part of its catalytic effect. Plotting the calculated free energy barrier heights of the mutant systems against the experimental tritium exchange rate produced a significant but imperfect correlation. It was concluded that reduction in rate observed for these mutants is a result of perturbations to the rates of both the Co-C bond breaking and hydrogen abstraction steps.

Given that the aim of achieving convergence and consistency with respect to the direction of modelling of the Co-C bond breaking has not been achieved, interpretation of these results in relation to experiment should be done tentatively, however they may still be able to provide some insight into the energetics of this process. The k_{cat} for turnover of glutamate by glutamate mutase is 5.8 s^{-1} at $22 \text{ }^{\circ}\text{C}$, corresponding to an activation free energy of 16.2 kcal/mol calculated with the Eyring equation.¹⁴³ This gives an upper limit to the free energy barrier for Co-C bond breaking. One-dimensional umbrella sampling of Co-C bond breaking in the backward direction gave a free energy barrier of 11 kcal/mol and a reaction energy of +4.5 kcal/mol. Assuming that this simulation is biased towards the dissociated state and therefore underestimates the reaction energy for Co-C bond breaking, it presumably also underestimates the barrier to Co-C bond breaking. This puts 11 kcal/mol as the lower limit to the barrier and +4.5 kcal/mol as the lower limit to the reaction energy for this step. Both of these may be lowered if the effects of zero-point energy were included, as the associated state should intuitively have higher zero-point energy than the transition and dissociated states due to the Co-C bond. One and

two-dimensional umbrella sampling produced clear minima associated with the dissociated adenosyl radical. Despite the reservations about the accuracy of these surfaces, this does provide some evidence that the adenosyl radical exists as an intermediate. The best estimate of the free energy barrier for hydrogen abstraction from umbrella sampling is 14.7 kcal/mol. Adding this to the lower limit for the reaction energy gives a combined free energy barrier for these steps of 19.2 kcal/mol. This is higher than the maximum barrier predicted from experiment. There is experimental evidence that quantum tunnelling significantly enhances the rate of hydrogen transfer in glutamate mutase.¹⁴⁴ If tunnelling and zero-point energy were included in the calculation of the hydrogen transfer step, the free energy barrier could feasibly be reduced by several kcal/mol. The reaction energy for the hydrogen transfer step calculated by umbrella sampling is around -4 kcal/mol. Combined with the lower limit for the reaction energy for the Co-C bond breaking step, this gives an overall reaction energy for these two steps of +0.5 kcal/mol. This is consistent with the experimentally observed accumulation of the glutamyl radical and consequent interpretation that the equilibrium constant for Co-C bond breaking is close to 1.^{132,145} Figure 41 shows a possible reaction profile based on these estimates.

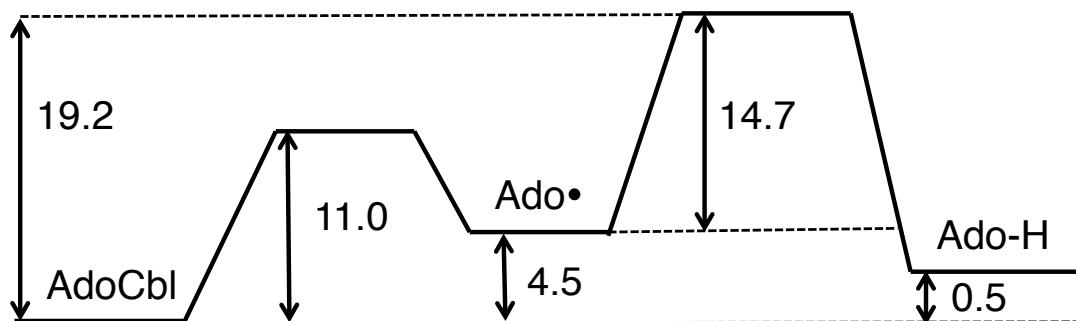


Figure 41 An estimated reaction profile for the Co-C bond breaking and hydrogen abstraction steps in glutamate mutase based on free energy calculations described in this work. All numbers are free energies in units of kcal/mol.

6.2. Further work

It is proposed that the structural restrictions imposed by the stochastic boundary conditions could be the cause of the failure to produce internally consistent converged

free energy profiles. To test and hopefully remedy this, the umbrella sampling simulations could be repeated with the full enzyme in periodic boundary conditions. This would require alterations to the EVB code in CHARMM as discussed in the previous chapter. If a reliable model of Co-C bond breaking was obtained, one could conceive of numerous simulations likely to lead to insight into adenosylcobalamin dependent catalysis. For example, Co-C bond breaking could be simulated in the absence of substrate to test the hypothesis that substrate binding facilitates this step. The charges on the adenosyl and/or the protein residues could be set to zero to test the theory that electrostatic stabilisation of the dissociated state drives catalysis. Free energy profiles could be calculated for the mutant systems to see if, when combined with those already calculated for the hydrogen abstraction steps, the experimentally observed trend in activity can be reproduced. The EVB potentials developed in this work should be trivially transferable to all other adenosylcobalamin dependent systems. Simulations could be performed on these other systems to investigate similarities and differences between them.

The two state EVB potentials for both steps could be combined in a three state model. The zeroth order iteration of this three state model would take the existing parameters exactly as they are and assume zero coupling between the associated state and the hydrogen abstraction state. This could even be extended to a four state model in which both hydrogens attached to the CG carbon are abstractable. With this three or four state model, the kinetic coupling between the Co-C bond breaking and hydrogen abstraction steps could be investigated.

It is clear from the present work that the reactions catalysed by adenosylcobalamin dependent enzymes involve numerous conformational changes, all of which will have associated barriers resulting in a large complicated network of interconnected minima. The group of David Wales in Cambridge has developed techniques for efficiently exploring complicated potential energy surfaces to find global minima and calculate rates for chemical processes. Judith Rommel, a research fellow working in the Wales group has interfaced the EVB potentials described in this work with the OPTIM code of the Wales group with the intention of exploring the conformational space of glutamate mutase to provide further insight into catalysis.

On the basis of experimental kinetic isotope effect (KIE) experiments, quantum tunnelling has been shown to occur to a significant but varying extent in several adenosylcobalamin dependent enzyme catalysed reactions as well as in solution. The

primary deuterium KIE was found to be 6, 56 and 42 in glutamate mutase, methylmalonyl-CoA mutase and adenosylcobalamin in ethylene glycol respectively.^{67,146,147} Whether hydrogen tunnelling is important for catalysis and why there is such variation in the experimental KIEs for apparently highly similar systems remains unclear. Other groups have calculated tunnelling rates in adenosylcobalamin dependent systems, but none have attempted to calculate KIEs for all of these systems using the same techniques in order to try to understand the difference between them.¹⁴⁸⁻¹⁵⁰ Ring polymer molecular dynamics (RPMD) is a fairly recently developed method that is capable of calculating highly accurate reaction rates accounting for quantum effects such as zero-point energy and tunnelling.^{4,151} As part of this PhD work (although not documented in this thesis), ring polymer molecular dynamics (RPMD) was implemented in CHARMM in collaboration with David Glowacki, Scott Habershon and Tom Markland. So far this implementation has been tested only on simple systems and requires the addition of some further functionality before quantum reaction rates can be calculated for enzyme systems. Once this work has been completed, RPMD could be used in conjunction with the EVB potentials described here to calculate KIEs in glutamate mutase and other adenosylcobalamin dependent systems.

Chapter 7. Introduction to heme dioxygenase enzymes

7.1. Introduction

7.1.1. Medical interest in the heme dioxygenases

The heme dioxygenases are a class of enzymes that catalyse the first step in the metabolism of L-tryptophan (L-Trp) via the kynurenine pathway, incorporating both atoms of molecular oxygen into L-tryptophan at the C2 and C3 positions to form N-formyl-L-kynurene (Figure 42). In humans, there are three known heme dioxygenases: indoleamine 2,3-dioxygenase (IDO), tryptophan 2,3-dioxygenase (TDO) and the more recently discovered indoleamine 2,3-dioxygenase 2 (IDO2). L-tryptophan is an essential amino acid for humans, meaning that it is indispensable for life but cannot be synthesised *de novo* in the body.¹⁵² 99% of all L-tryptophan metabolism occurs via the kynurenine pathway, so the local and global levels of L-tryptophan in the body are strongly linked to heme dioxygenase activity and expression.¹⁵³

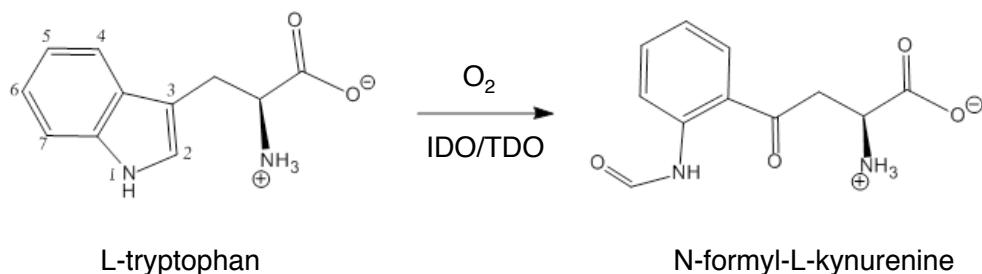


Figure 42 Reaction catalysed by the heme dioxygenase enzymes.

The main biological function of TDO is thought to be the regulation of the supply of the various metabolites of the kynurenine pathway such as nicotinamide adenine

dinucleotide (NAD).¹⁵⁴ The importance of IDO as an immunoregulatory enzyme was revealed when it was discovered that inhibition of IDO in pregnant mice led to foetal rejection.¹⁵⁵ IDO was found to be involved in the suppression of the immune response towards the foetus, which would otherwise be considered a pathogenic foreign body. Little is known about the function of IDO2 due its recent discovery, although it too appears to have a role in immunoregulation.¹⁵⁶

IDO and TDO have been implicated in immune escape mechanisms employed by tumours and so are both considered potential targets for cancer treatment.¹⁵⁷ The IDO inhibitor 1-Me-D-Trp has recently been the subject of clinical trials.^{158,159} An important property of any potential heme dioxygenase inhibitor is that it should be highly specific to one enzyme. Due the biological importance of these enzymes, it is of considerable interest to study their reactivity, substrate binding and inhibition, particularly with a view to gaining new insight into the similarities and differences between them.

7.1.2. Structure and substrate binding

Significant advances in the understanding of substrate binding and reactivity in the heme dioxygenases came with the solving of several structures by x-ray crystallography. First, the structures of heme-bound human IDO (hIDO) in complex with the inhibitors 4-phenylimidazole and cyanide were solved (pdb codes 2D0T and 2D0U)¹⁶⁰, closely followed by the structures of heme-bound bacterial TDO (XcTDO) with and without L-Trp bound in the active site (pdb codes 2NW8 and 2NW7).¹⁶¹

These structures revealed that despite a very low level of sequence identity between hIDO and XcTDO, their active site structures are remarkably similar (Figure 43). The crystal structure of TDO with substrate bound identified several key binding interactions. A salt bridge is formed between Arg117 and the carboxylate group of the substrate. This arginine residue is conserved across all known heme dioxygenases and has been shown by several mutagenesis studies to be essential to activity.^{162,163} The heme 7-propionate group and Thr254 both interact with the NH₃⁺ group on the substrate.

Comparison of the substrate free and substrate bound structures of XcTDO suggest that a significant change in active site structure accompanies substrate binding (Figure 44).

In the substrate free structure, Arg117 points out towards the solvent and the section containing residues 252 to 256 is missing from the structure due to disordered electron density. In the substrate bound structure, Arg117 has moved towards the interior of the active site to form a salt bridge with the substrate. Thr254 (which is disordered in the substrate free structure) forms a hydrogen bond with the substrate. The indole N-H hydrogen on the substrate forms a hydrogen bond with His55.

Comparison of the substrate free XcTDO structure and the hIDO structure (in which inhibitors lacking any carboxylate or NH_3^+ groups are bound in the active site) reveals several similarities. In hIDO, Arg231 (equivalent to XcTDO's Arg117) points out of the active site towards the solvent. The hIDO structure is also missing a section (residues 261 to 279) in a location analogous to that which is missing from the substrate free XcTDO structure. A notable difference is that the missing section is further away from the substrate and heme in hIDO than in XcTDO. Due to the similarities in active site structure and reactivity between XcTDO and IDO, it can be expected that the conformational changes and substrate binding interactions that are observed in XcTDO also occur in hIDO.

As there is no O_2 molecule bound to any of the crystal structures, the nature of possible interactions of the protein and substrate with O_2 was unclear. It had been suggested that the substrate indole N-H hydrogen bonded directly to the terminal oxygen.¹⁶⁴ A more recent study using ENDOR spectroscopy gave strong evidence that the NH_3^+ group of the substrate, not the indole N-H group, hydrogen bonds directly to the terminal oxygen.¹⁶⁵ This has important implications for the reaction mechanism as it is yet another factor undermining the suggestion that the indole N-H proton is abstracted by oxygen in the first step of the reaction (as will be discussed later).

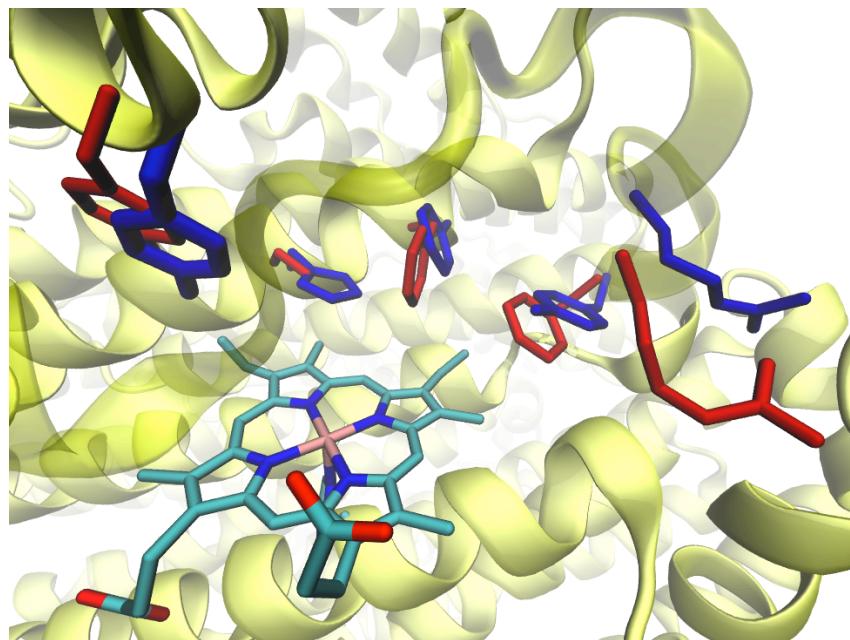


Figure 43 Superposition of the active sites of hIDO (red) and XcTDO (blue). The 4-phenylimidazole bound hIDO structure (pdb code 2D0T) and the substrate free XcTDO structure (2NW7) are shown.

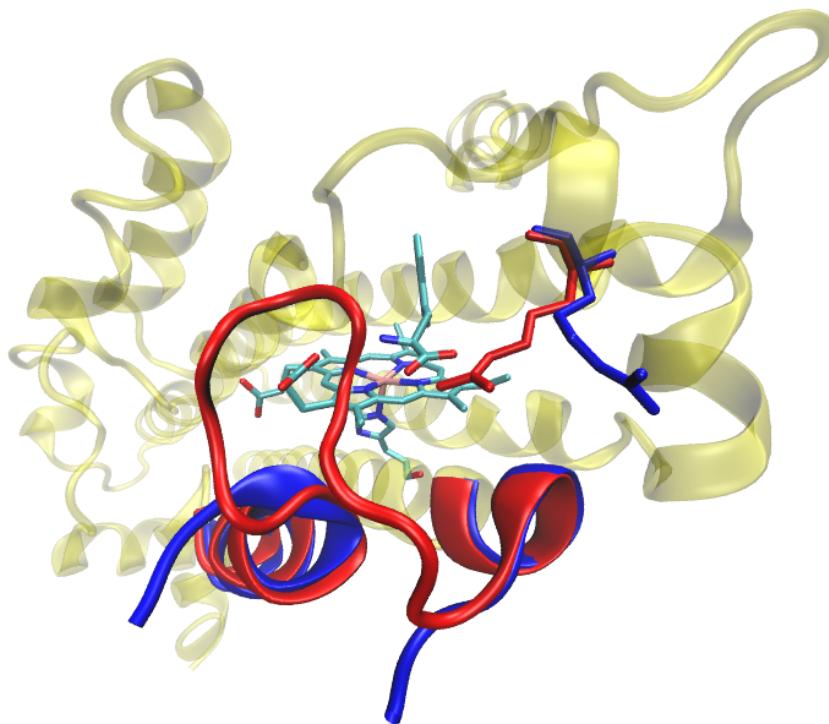


Figure 44 Superposition of two crystal structures of XcTDO. The structure in red was crystallised with L-Trp bound (pdb code 2NW8). The structure in blue was crystallised without substrate bound (pdb code 2NW7). The structures are very similar globally but differ significantly in the structure of an active site loop section and the position of Arg117 (shown in blue and red).

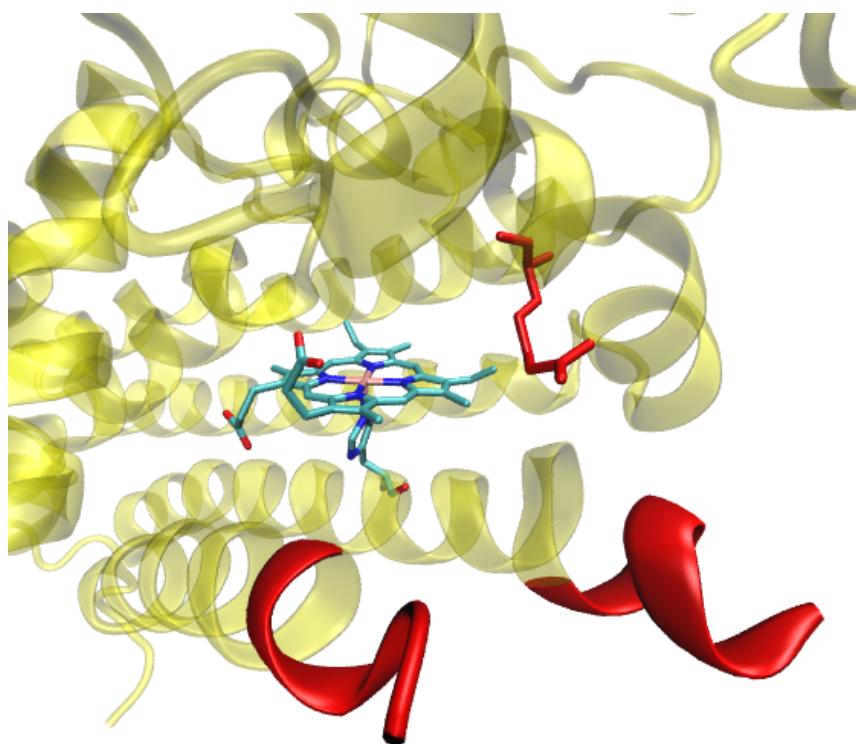


Figure 45 Crystal structure of inhibitor bound hIDO (pdb code 2D0T). Shown in red are Arg231 (analogous to Arg117 in XcTDO) and the start of the missing loop section.

7.1.3. Substrate specificity

TDO is highly specific to L-Trp and a small number of very similar analogues such as 5-F-Trp, and 6-F-Trp.¹⁶¹ IDO is active with wider range of L-Trp analogues such as D-Trp, 1-Me-L-Trp, serotonin, tryptamine and 5-OH-Trp among others.¹⁶⁶ Kinetic experiments have been performed on many of these substrates, however this study will focus on three substrates only: L-Trp, D-Trp and 1-Me-L-Trp (Figure 46). D-Trp and L-Trp were chosen because how biological systems in general differentiate between subtly different L- and D- enantiomers is a question of fundamental interest.¹⁶⁷ 1-Me-L-Trp was chosen for study because of the fairly recent discovery that it is a slow substrate for hIDO (and mutant XcTDO enzymes) which has implications for proposed reaction mechanisms (discussed below).¹⁶⁸ Table 15 summarises the experimental kinetic data for these substrates with hIDO, XcTDO and the H55A XcTDO mutant.

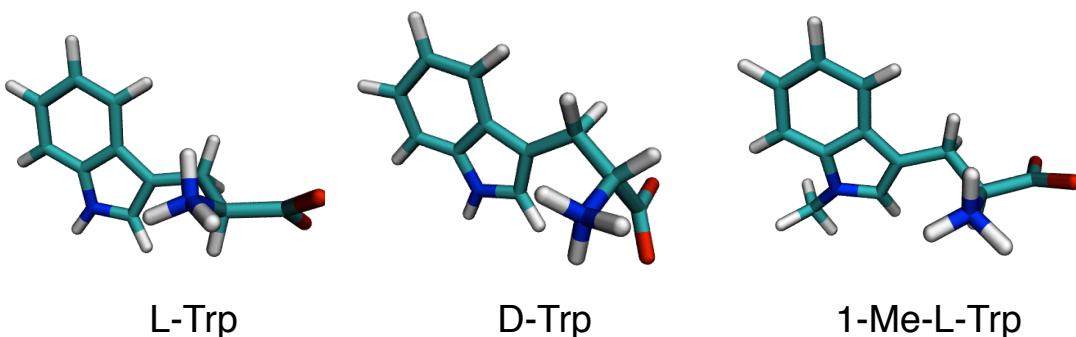


Figure 46 The three substrates that are the focus of this investigation into reactivity and binding in the heme dioxygenases.

Table 15 Kinetic data for reaction of L-Trp, D-Trp and 1-Me-L-Trp with hIDO, XcTDO and the H55A XcTDO mutant. Data originates from a¹⁶⁹ b¹⁶⁸ c¹⁷⁰ d¹⁶¹. The K_M value for TDO with D-Trp is actually an inhibitor constant K_i. Experiments were unable to detect any activity with 1-Me-L-Trp in TDO.

	k _{cat} (s ⁻¹)	K _M (μM)	ref.
IDO Trp	1.4	7	a
IDO 1-Me-L-Trp	0.027	150	b
IDO D-Trp	3.93	1570	a
TDO L-Trp	19.5	144	c
TDO 1-Me-L-Trp	-	-	b
TDO D-Trp	0	16500	d
TDO-H55A L-Trp	2.86	133	c
TDO-H55A 1-Me-L-Trp	0.048	59	b

7.2. Reaction mechanism

Early reaction mechanisms proposed for this enzyme catalysed reaction involved abstraction of the indole N-H proton by either an active site base or by the heme-bound dioxygen as the initial step (Figure 47a,b).^{160,171} This was considered essential because experimental evidence indicated that 1-methyl-L-tryptophan (1-Me-L-Trp), which does not have an indole N-H hydrogen, is a competitive inhibitor of both enzymes.¹⁷² Subsequent steps were suggested to be nucleophilic attack on the distal oxygen at the C3 position followed by product formation via either a dioxetane intermediate or a Criegee rearrangement. More recent experimental work has shown 1-Me-L-Trp to in fact be a slow substrate for IDO and mutant TDO systems calling into question all mechanisms involving a proton abstraction initial step.¹⁶⁸

A density functional theory (DFT) study on a model system (comprised of a truncated oxy-heme group with imidazole representing the axial histidine ligand and 3-methyl-indole representing the substrate) was the first computational study of the heme dioxygenase reaction mechanism.¹⁷³ Whereas previous mechanisms had assumed the oxy-heme complex to be in the closed shell ferrous dioxygen state Fe(II)-O₂, this study found the open shell singlet state to be significantly lower in energy, which (based on Mulliken spin population analysis) was considered to a ferric superoxide complex Fe(III)-O₂•-. The results of the DFT study showed that mechanisms involving indole N-H proton abstraction by dioxygen are very unfavourable. Newly proposed mechanisms starting with electrophilic addition from the Fe(II)-O₂ (Figure 47c) closed shell singlet state or radical addition from the Fe(III)-O₂•- triplet or open shell singlet states gave more reasonable barriers. Addition to the C2 carbon was found to be lower in energy than to the C3 carbon. Following addition, the Criegee rearrangement pathway was shown to be very unfavourable. Homolytic cleavage of the O-O bond leading to the formation of an epoxide and the high valent Fe(IV)=O species, compound II, was considered to be feasible though higher in energy than the dioxetane pathway. Shortly afterwards, compound II was identified by resonance Raman spectroscopy during IDO turnover (although not during TDO turnover).¹⁷⁴

The solution of crystal structures for hIDO and bacterial XcTDO provided a starting point for hybrid quantum mechanical/molecular mechanical (QM/MM) calculations, which have been performed by three different groups. One studied both IDO and TDO (Capece et al.)¹⁷⁵ while the others performed calculations on TDO only (Wallrapp and Guallar)¹⁷⁶, (Chung et al.).¹⁷⁷ Each study used different software packages and levels of theory in their calculations, but the general approach was the same: using molecular mechanics force fields, several nanoseconds of molecular dynamics simulations were run from which structures were chosen as starting points to explore the DFT QM/MM potential energy surfaces.

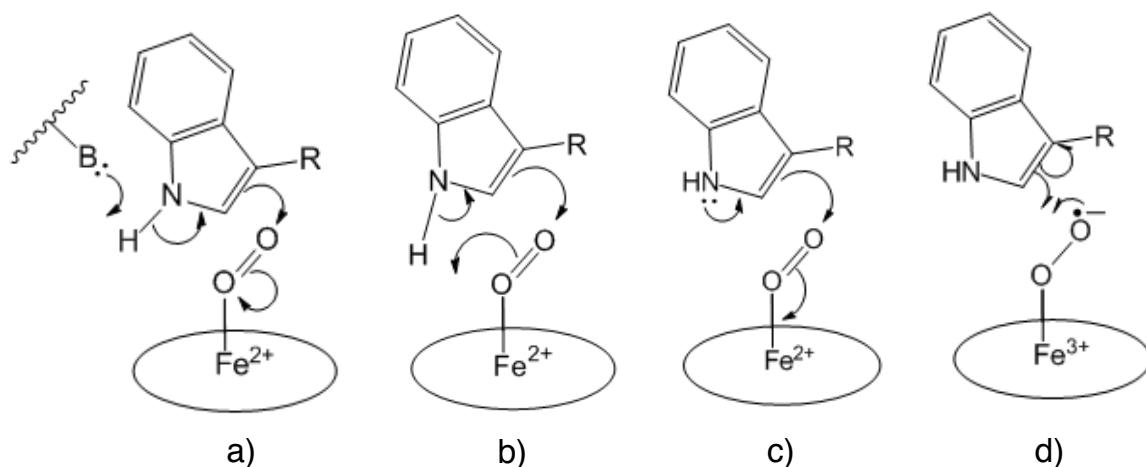


Figure 47 First steps for a number of proposed reaction mechanisms. a) and b) involve abstraction of the indole N-H proton from either an active site base or dioxygen.¹⁷¹ c) involves electrophilic addition in the ferrous-oxy state.¹⁶⁸ d) involves radical addition in the ferric-superoxide state.¹⁷⁷

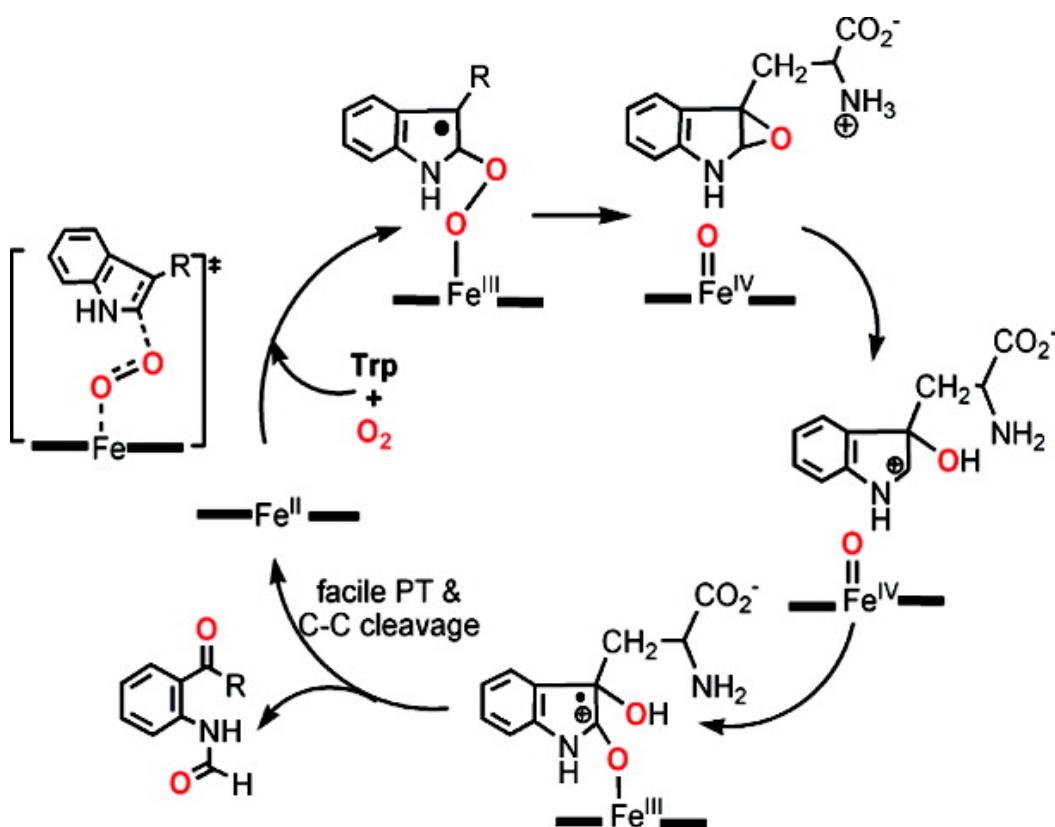


Figure 48 The mechanism proposed as a result of an ONIOM study on TDO (figure reproduced exactly from the paper).¹⁷⁷ The first step involves radical addition in the ferric-superoxide state to form an alkylperoxo species. This is followed by homolytic cleavage of the O-O bond to form an epoxide and compound II (rate limiting step).

These QM/MM studies all concluded that abstraction of the indole N-H proton does not occur because no minimum is observed along any of the proton transfer reaction coordinates investigated. This is in agreement with experimental results that showed that the only candidates for the proposed active site bases (His55 in TDO and Ser167 in IDO) are not essential to activity and that 1-Me-L-Trp is a slow substrate for IDO and mutant TDO enzymes.^{168,178} Capece et al. and Chung et al. investigated the dioxetane pathway and found it to be infeasible in enzyme.^{175,177}

In all these QM/MM studies, the open shell singlet Fe(III)-O₂•- state was found to be the lowest energy state of the Michaelis complex and radical addition to the C2 carbon was found to occur followed by the formation of an epoxide and compound II. Oxygen addition to the C3 carbon did not result in a minimum. Epoxide and compound II formation was found to be the rate-determining step in all cases, with energy barriers ranging from 8 to 20 kcal/mol. One notable difference between the studies is that the alkylperoxy species resulting from radical addition was identified as being a transition state in the study by Capece et al. whereas the others found this species to be an intermediate.

Chung et al. calculated pathways in both the singlet and triplet states for the first two steps and found that the surfaces cross before the transition state to compound II and epoxide formation where the triplet surface becomes lower in energy. The rest of the reaction profile is given on the triplet surface only, although how or where a change in spin state may occur is not discussed.

The calculations of Capece et al. suggest that IDO and TDO follow the same reaction mechanism, which aside from the concerted versus stepwise nature of the epoxide and compound II formation is in qualitative agreement with the mechanism proposed by Chung et al. for TDO. Based on the results of these studies, the leading proposed reaction mechanism for the first two steps is that shown in Figure 48. All of these studies calculate full pathways to product formation, with variations in mechanism between them. As epoxide and compound II formation was found to be cleanly rate limiting, these further steps should not influence substrate specificity and so are not considered in this study.

Chapter 8. MD study of substrate binding in IDO and TDO

8.1. Introduction

In recent years, significant progress has been made in the understanding of substrate binding and reaction mechanism in the heme dioxygenases. As discussed in the previous chapter, this is as a result of x-ray crystallography, spectroscopy, mutagenesis and kinetics studies, as well as several insightful computational studies. The main aim of the present work is to investigate the origin of the differences between IDO and TDO with regards to their reaction and binding specificity. In particular, the binding and reactivity of these enzymes with L-Trp, D-Trp and 1-Me-L-Trp will be investigated. In addition, further detailed insight into the mechanism of heme dioxygenase catalysis is sought.

The crystal structure of XcTDO with the natural substrate (L-Trp) bound has provided significant insight into substrate binding in the heme dioxygenases. However, how active site structure and interactions might change with addition of O₂ or the binding of a different substrate is unclear. The available crystal structures for hIDO are bound to inhibitors, which are not particularly similar to L-Trp. While it is reasonable to assume that IDO and TDO will bind L-Trp in a similar manner, they cannot bind in exactly the same way because IDO lacks residues homologous to His55 and Thr254 (which form hydrogen bonds with the substrate in XcTDO). Building on the experimental structural evidence, molecular modelling may be able to provide useful additional insight. Due to the considerable interest from the medical community in finding inhibitors for the heme dioxygenases, numerous groups have simulated the binding of various different ligands or ligand fragments to IDO and TDO using docking methods.^{179–182} Whilst these studies have provided some useful insight and leads for drug design, it is known that results of docking simulations (especially when not performed in conjunction with MD simulations) can be misleading.¹⁸³ This is in large part because enzyme active sites, being flexible and dynamic, will adapt to the presence of different ligands in a way that cannot

be easily modelled with docking tools, even if a certain amount of flexibility is accounted for. This is likely to be a major effect in substrate binding to the heme dioxygenases, as it is known from the crystal structures of XcTDO that significant conformational change accompanies substrate binding.¹⁸⁴

The main aim of the work described in this chapter is to provide insight into the differences in substrate binding specificity between IDO and TDO by modelling the structure and dynamics of L-Trp, D-Trp and 1-Me-L-Trp bound to IDO, TDO and the H55A TDO mutant and by calculating the associated binding free energies. In the following chapter, the reactivity of these systems will be modelled with QM/MM adiabatic mapping calculations. In order to obtain meaningful results from adiabatic mapping, it is extremely important to start from a range of structures which accurately represent the Michaelis complex.⁴⁹ A secondary purpose for these MD simulations is to generate an equilibrium ensemble of structures for each system for use in QM/MM calculations.

Note: Although L-Tryptophan binding to IDO was studied using MD simulations in the author's final year undergraduate project, all simulations and analysis presented here were undertaken following the start of this PhD project.

8.2. Computational details

8.2.1. TDO specific system preparations

The x-ray crystal structure of TDO from *Xanthomonas campestris* (XcTDO) in complex with L-Tryptophan (PDB code 2NW8) was used as the starting point for all calculations on TDO.¹⁸⁵ Dioxygen was added to the each heme using the Molefacture VMD plugin with an estimated geometry based on x-ray structures of oxy-myoglobin from the protein data bank.^{127,186,187} L-Tryptophan is present in both active sites A and B (although the binding pose and active site structure differ between these sites). The crystallographic waters with residue number 402 were removed from each active site due to steric clashes with dioxygen and the substrate. Two crystallographic Mn²⁺ ions were removed from the surface of the protein far from the active sites.

Coordinates for chains A and B only were given (as well as their associated heme groups and substrates). Symmetry operations were performed in Swiss-Pdb Viewer to complete the biologically active homotetrameric assembly.¹⁸⁸

Residues 251 to 256 from chain B (and consequently chain D also) were missing from the x-ray structure. These residues form part of the loop which is thought to block the active site from solvent in the closed form of the enzyme, so it was considered important to generate coordinates for this region.¹⁸⁵ To complete the structure, chain A was superimposed onto chain B. The coordinates for residues 248 to 260 of the superimposed chain A were used as the coordinates for those residues in chain B. All superpositioning described in the setup of the heme dioxygenase enzyme systems was performed using Swiss-Pdb Viewer.¹⁸⁸ A short minimisation was then performed to reduce steric clashes using the GROMOS 43B1 force field within Swiss-Pdb viewer.¹⁸⁹

Coordinates were also missing for the first 18 residues of each chain and for residues 285 to 306 of chain A and C and residues 291 to 306 of chains B and D. As these residues are at the ends of the sequence, are probably largely unstructured and are not particularly close to the active site they were simply left out of the model.

The H55A TDO mutant structure was generated by manually editing the pdb file, replacing the imidazole of histidine with a hydrogen.

8.2.2. IDO specific system preparations

The x-ray crystal structure of human IDO in complex with the inhibitor 4-phenylimidazole (pdb code 2D0T) was used as the starting point for all simulations of IDO.¹⁶⁰ A section analogous to that which was missing from chain B of the TDO structure was also missing from the IDO crystal structure (residues 361 to 379). A model for this section was generated by analogy to sections of other proteins with similar sequence and known 3D structure using the loop building tool in Swiss-Pdb Viewer.¹⁸⁸ An investigation of several models for this loop section was undertaken as part of the author's final year undergraduate project.¹⁹⁰ A full description of that work will not be reproduced here, but the model of the missing section studied in that work believed to be most reasonable was used in all MD simulations of IDO reported here. The first 12 residues were also not present in the crystal structure but were not modelled for reasons given above for the analogous region of the TDO structure.

The inhibitor and two molecules of 2-(N-cyclohexylamine) (which form salt bridges with the heme 7-propionate moiety and with Arg231) were removed from the system. L-Tryptophan was added to the structure by superimposing the L-Tryptophan bound TDO active site A onto the IDO active site and copying the L-Tryptophan coordinates. Dioxygen was added in the same manner as with TDO.

8.2.3. Setup procedure for all systems

The pKas of ionisable residues were calculated with the web server PROPKA.¹²⁴ Protonation states of histidine residues were determined based on these calculated pKas and by visual inspection of implied hydrogen bonds. The calculated pKas showed all other ionisable residues to be in their standard protonation states at pH 7. Some Asn and Gln residues were flipped (the positions of OD1 and ND2 exchanged) to remove steric clashes as calculated with the structure validation programme Molprobity.¹²⁵

D-Tryptophan and 1-Me-L-Tryptophan systems were generated by manipulating the L-Tryptophan bound systems using the Molefactory VMD plugin.¹²⁷

The psfgen plugin for VMD was used to solvate, add ions and generate the topology files (psf) for use with NAMD.¹²⁷ Each system was solvated in a box of pre-equilibrated CHARMM TIP3P water molecules with at least 12 Å between the edge of the box and any protein atoms. Each system was neutralised by adding sodium and chloride ions at a concentration of 0.05 mol/L.

8.2.4. MD simulations

The CHARMM 27 all atom force field was used for all simulations.¹⁹¹ Additions to this force field were made previously within the group to enable modelling of dioxygen bound to ferrous heme.¹⁹² These parameters were not thoroughly tested for use in long scale MD simulations (as this was not their original purpose). Consequently there are some notable deficiencies, which have come to light over the course of this study. These should ideally be addressed before these parameters are used for further long scale MD of oxyheme systems. Dioxygen is known to bind to heme in a bent orientation, however there is no Fe-O-O angle term in these parameters. There are also no O-O-Fe-N* dihedral terms (where N* represents any of the Fe coordinating nitrogens). In the first iteration of

these parameters, there was no partial charge on either of the heme bound dioxygen atoms. This was adjusted so that there was a -0.21 and +0.21 charge on the terminal and proximal oxygens respectively. These charges are rough estimates based on Mulliken population analysis oxyheme using the same heme model and DFT methodology as is described in section 9.2.3.

All minimisations and MD simulations were performed in NAMD using the same procedure described in previous work by the Mulholland group.^{193,194} Periodic boundary conditions were used. The Particle-Mesh-Ewald method was used for electrostatic interactions with a grid density of approximately 1 per Å. Van der Waals' interactions were scaled using a switching function between 12 and 10 Å. The conjugate gradient method was used for all minimisations. Leapfrog integration with a 2fs time step was used for all MD simulations. The SHAKE algorithm was applied to constrain all bonds containing hydrogen atoms.¹⁹⁵

The first stage was the removal of bad contacts due to the guessed hydrogen atom positioning. The system was minimised for 3000 steps with all heavy atoms held fixed.

The next stage was to heat and equilibrate the solvent, with all hydrogen and all solvent atoms free to move, but all heavy solute atoms still fixed. The system was minimised for 6000 steps. 5ps of MD was performed using temperature reassignment to heat the solvent from 50K to 300K in 50K increments every 0.5ps. 15ps of NVT MD was performed using Langevin dynamics with a Langevin damping coefficient of 5ps⁻¹.

For the final stage of equilibration, the whole system was minimised then heated and equilibrated with all atoms free to move. The system was minimised for 10000 steps. 20ps of MD was performed using temperature reassignment to heat the whole system from 50K to 300K in 50K increments every 2ps. 180ps of NPT MD was performed using Langevin dynamics with only the heavy atoms coupled to an external heat bath, again with a Langevin damping coefficient of 5ps⁻¹. Constant pressure was maintained using the Nosé-Hoover Langevin piston pressure method with a piston decay time scale of 50fs and a piston oscillation period of 100fs.^{196,197}

The MD production phase was performed in the NPT ensemble using the same setup as the end of the equilibration but with a piston oscillation period of 200fs and a piston decay time scale of 100fs. The entire equilibration and production procedure was repeated for each separate run with different velocities randomly assigned during the heating stages. 20ns of production phase MD was performed for each production run.

One production run of IDO with L-Trp was extended to 50ns. A summary of all MD simulations performed is given in Table 16.

Table 16 Summary of all MD simulations in this study. Note that as there are 4 active sites in TDO, each production run provides 4 times more data (for the purposes of this study) than each IDO production run.

enzyme	substrate	amount of MD
IDO	L-Trp	4 x 20 ns, 1 x 50 ns
IDO	1-Me-L-Trp	5 x 20 ns
IDO	D-Trp	5 x 20 ns
IDO	no substrate	1 x 20 ns
TDO	L-Trp	2 x 20 ns
TDO	1-Me-L-Trp	2 x 20 ns
TDO	D-Trp	2 x 20 ns
TDO H55A	L-Trp	2 x 20 ns
TDO H55A	1-Me-L-Trp	2 x 20 ns

8.2.5. Analysis tools

Coordinates were saved every 2ps to generate trajectory files for analysis. Unless otherwise stated, all analysis presented here was done on the section of each trajectory between 10 and 20ns.

Root mean square deviation (RMSD), hydrogen bonding, salt bridge and all other geometric analysis was performed using VMD in scripting mode.¹²⁷ To calculate the RMSDs presented here, the structures from each frame in a trajectory were rotated and translated so as to give the minimum RMSD of the coordinates of the specified atoms from the coordinates of those atoms in the relevant crystal structure. The specified atoms for the three different types of RMSD calculations were: 1) C α backbone atoms, 2) all C α backbone atoms excluding the loop region that was missing from the crystal structure for IDO and 3) the C α backbone atoms for all residues with at least one atom within 6 Å of the heme group in relevant crystal structure.

Hydrogen bonding analysis was performed with the hbonds module of VMD. A hydrogen bond was considered present in a particular structure if the donor-acceptor distance was less than 3.5 Å and the donor-hydrogen-acceptor angle was less than 35°. A

salt bridge was considered present if the distance between the specified heavy atoms was less than 4.5 Å. These distances are illustrated in Figure 53.

MM/PBSA and MM/GBSA binding free energy calculations were performed using the MMPBSA.py module of AmberTools14, calling the sander module of AmberTools14.¹⁹⁸ For the theoretical background and a critical assessment of performance of MM/PBSA and MM/GBSA see the following references.^{199,200} The single trajectory protocol was used. As input, MMPBSA.py requires sander compatible prmtop (parameter and topology) files for the ligand (L-Trp or analogues), receptor (protein + heme + O₂), complex (ligand + receptor), solvated complex (complex + water + ions) and a trajectory file (of the solvated complex). As the MD simulations were originally run in NAMD with a psfgen generated topology file, some conversion was necessary to generate the prmtop files. The CHAMBER module of AmberTools14 was used to convert CHARMM force field psf (protein structure) and parameter files for the relevant subsystems into sander compatible prmtop files.²⁰¹ At the time of simulation, this was only possible starting from CHARMM generated psf files and not with VMD psfgen generated psf files (although a recent AmberTools patch has apparently addressed this). CHARMM version 36a1 was used to generate the necessary psf files for conversion into prmtop files by CHAMBER.²⁰² The vibrational entropic contribution to the binding free energy was not calculated as this type of calculation was not possible with CHAMBER generated prmtop files in the version of AmberTools14 used in this work. The neglect of this contribution is not expected to make a large difference to the relative binding free energies of the various substrates for the same enzyme because of the high degree of similarity between the substrate structures.

Structures were taken at evenly spaced intervals from the 10th to 20th nanosecond of each trajectory for binding free energy calculations. To test for convergence with respect to the number of structures used, MM/PBSA binding free energy calculations were performed on each of the trajectories from the 5 IDO with L-Trp simulations with 60, 120, 300 and 600 structures. The calculated binding free energies did not appear to be converging to a single value in all cases as the number of structures was increased. However, for each trajectory, the range of binding free energies calculated with different numbers of structures was less than 1 kcal/mol in all cases. Consequently, any number of structures greater than 60 was considered sufficient to achieve convergence of binding free energy with respect to the number of structures used. Overall convergence of binding

free energies between difference production runs of the same system was not achieved as will be discussed in more detail later.

For IDO systems, 120 structures were used from each trajectory and for TDO systems only 75 structures were used as the calculations took longer due to the larger system size. Separate binding free energy calculations were done for each active site of each TDO system, with the other active sites considered part of the receptor.

8.3. Results and discussion

8.3.1. Structural equilibration

Plots of the RMSD of backbone positions were visualised for all MD simulations. RMSDs are reported in Figure 49 for one simulation of IDO with L-Trp (10th to 50th ns) and of one simulation of TDO with L-Trp (other plots are not shown as these plots are fairly typical of all systems). An in depth interpretation of the RMSDs will not be attempted here but it is worth noting that the RMSDs shown here continuously rise throughout the course of the simulations, so the simulations cannot be said to have reached a stable equilibrium ensemble of structures. It is especially worth noting that the RMSDs for IDO with L-Trp continue to rise beyond 20 ns, which is the length of most of the MD simulations reported here. In most cases, the RMSDs of the active site regions are fairly low, which bodes well for the reliability of the properties reported here as this work is primarily concerned with active site structure, substrate binding and reactivity.

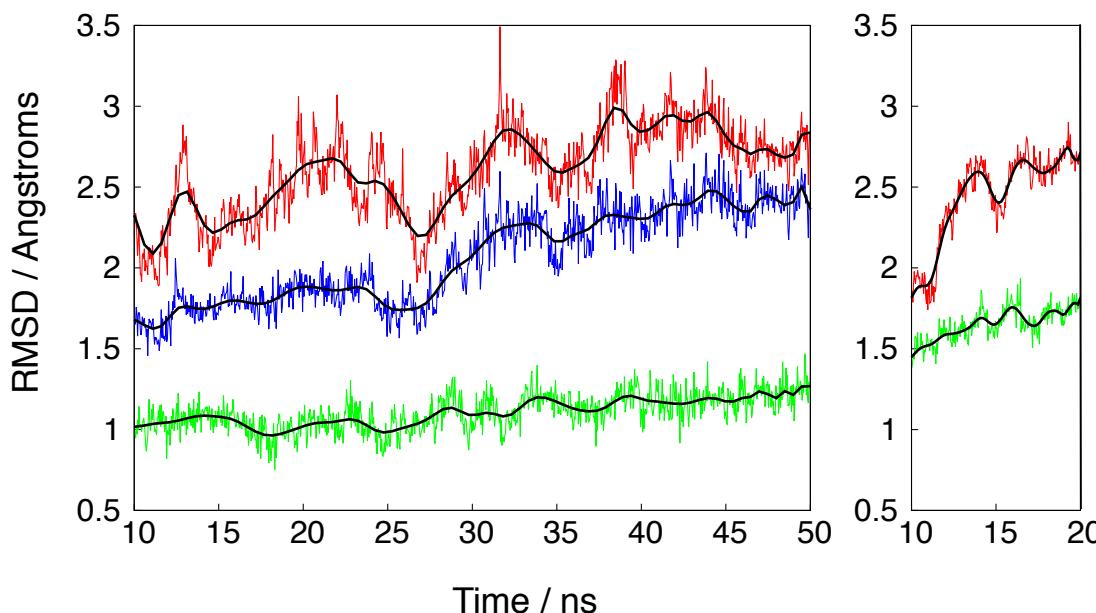


Figure 49 RMSD plots of single production runs of IDO with L-Trp (left) and TDO with L-Trp (right) with reference to their respective crystal structures. The plots shown are of the RMSDs of all C α backbone atoms (red), C α backbone atoms excluding the loop region that was missing from the crystal structure for IDO (blue) and C α backbone atoms for all residues with at least one atom within 6 Å of the heme group in the relevant crystal structure (green).

8.3.2. Binding free energy calculations

Binding free energies were calculated with MM/PBSA and with MM/GBSA for each system. Figure 50 compares the binding free energies calculated with each method. MM/GBSA on average gives a binding free energy 4 kcal/mol lower than MM/PBSA. There is a high degree of agreement between the two methods. Given this and the fact that MM/PBSA is a less approximate method, only the MM/PBSA results will be interpreted here.

Binding free energies were calculated by running MM/PBSA calculations on each active site separately using structures taken from individual production runs. This produced a set of 5 and 8 calculated binding free energies for each IDO and TDO system respectively. Table 17 summarises the results of the binding free energy calculation for all systems studied. Before attempting to interpret the results, it is important to note that these calculations have clearly not converged to a satisfactory degree. This is indicated by the large standard deviations of the calculated binding free energies for each system.

relative to the difference in calculated (and experimental) binding free energies between systems. As mentioned in the computational details, all of the binding free energies calculated from a given active site from a given MD trajectory were converged with respect to the number of structures used. Therefore, the lack of convergence observed here must be due to variations in the ensembles of structures between separate active sites of separate MD simulations. Given this, the MD simulations plus MM/PBSA protocol reported here (in its current form at least) cannot be said to be an accurate tool for predicting the relative binding free energies of these systems. This failure is not unprecedented, as the performance of MM/PBSA in ranking ligands by binding free energy is known to be highly system dependent and can range from reasonably good to very poor.^{200,203,204}

Despite the reservations about the reliability of the final predictions of these calculations, there may still be some useful insight to be gained by analysing the results. Figure 51 (top) plots the mean calculated binding free energies of each system against the logarithm of the relevant experimental Michaelis constants (K_M). Note that the K_M given for TDO with D-Trp is actually a inhibitor constant (K_i). This is because K_M s cannot be determined if the system does not turn over at a detectable rate as was the case with TDO with D-Trp and 1-Me-L-Trp.^{168,184} No K_i could be found in the literature for TDO with 1-Me-L-Trp, although it should be possible to determine this value from experiment.

There appears to be some correlation between the mean calculated binding free energies and experimental Michaelis constants (K_M). However, given the small data set, the poor convergence just mentioned and the severely anomalous position of the data point for L-Trp in IDO, it is possible that this correlation is largely spurious. While it appears that the mean values for the calculated binding free energies are unreliable, it seems reasonable to assume that the lowest value calculated from an individual production run for each system will be the closest to the hypothetical converged binding free energy. Consequently, the ensemble of structures associated with this value should be most representative of the equilibrium state of the system. The reasoning behind this argument is that any system at equilibrium will spend more time in states of lower rather than higher free energy and that the simulations which produced high calculated binding free energies may have been trapped in metastable ensembles of structures that are not representative of the equilibrium ensembles. Using this logic, it was hoped that some consistency with experiment could be observed by looking only at the lowest calculated binding free energies for each system.

Figure 51 (bottom) plots the lowest calculated binding free energies for each system against the log of the experimental K_{MS} . While there is no apparent overall correlation, a trend does emerge if the systems are split into 3 groups: the IDO systems, the TDO systems which turn over (TDO with L-Trp, H55A TDO with L-Trp and H55A TDO with 1-Me-L-Trp) and the TDO systems which do not turn over (TDO with D-Trp and 1-Me-L-Trp). There is clearly no correlation among the IDO systems, which all have lowest values for the calculated free energies of between -32 to -33 kcal/mol despite a difference of three orders of magnitude in their experimental K_{MS} . MM/PBSA is known to be a poor method for calculating absolute binding free energies, so it is only reasonable to compare relative binding free energies of similar systems.²⁰⁵ Although IDO and TDO do share a high level of similarity, the two enzymes may be too different for their calculated binding free energies to be comparable. This system dissimilarity may explain the clustering of the lowest calculated binding free energies for the IDO systems at around 15 kcal/mol higher than those calculated for the active TDO systems.

Looking at the lowest values of the calculated binding free energies for the TDO systems alone, the results can be split into two groups. The free energies for the TDO systems which turn over are tightly clustered together, which is consistent with the relatively tight clustering of their K_{MS} . The TDO systems which do not turn over have significantly higher values (41 kcal/mol and 33 kcal/mol for TDO with 1-Me-L-Trp and D-Trp respectively). It is known that D-Trp does not bind favourably to TDO from the value of its experimental K_i . Without a K_i for TDO with 1-Me-L-Trp, it is unclear from experiment whether this system does not turn over due to poor binding, a high barrier or both. The lowest value for the calculated binding free energy for this system is consistent with the hypothesis that it binds considerably less well than the TDO systems which do turn over.

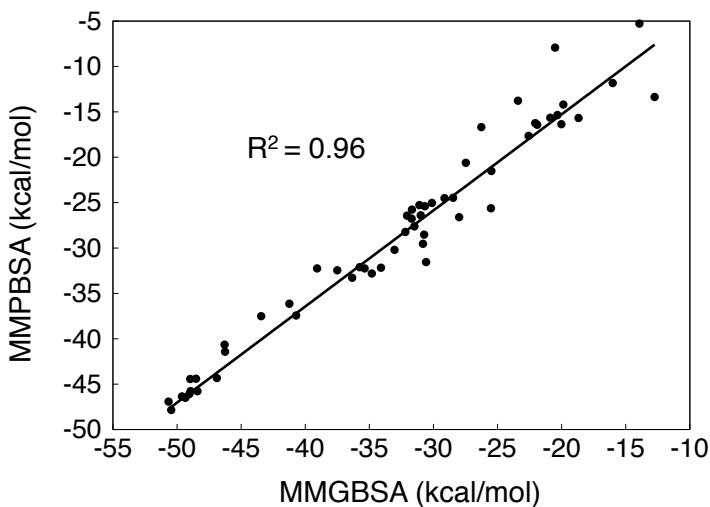


Figure 50 Binding free energies for each production run of each system as calculated with MM/PBSA plotted against those calculated with MM/GBSA.

Table 17 A single value for the binding free energy was calculated for each production run of each system. Presented here are the minimum, maximum, mean and standard deviation of these values for each system as calculated with MM/PBSA. The experimental K_M is given except where it was not available (as indicated with a “-”). The ref. column the source of the experimental data: a¹⁶⁹, b¹⁶⁸, c¹⁷⁰ and d¹⁸⁴. The K_M value for TDO with D-Trp is actually an inhibitor constant K_i .

	min. ΔG_{bind} (kcal/mol)	max. ΔG_{bind} (kcal/mol)	mean ΔG_{bind} (kcal/mol)	SD ΔG_{bind} (kcal/mol)	exp. K_M (μM)	exp. k_{cat} (s^{-1})	ref.
IDO Trp	-32	-14	-19	8	7	1.4	a
IDO 1-Me-L-Trp	-33	-24	-30	4	150	0.027	b
IDO D-Trp	-32	-16	-26	6	1570	0.93	a
TDO L-Trp	-47	-25	-36	11	144	19.5	c
TDO 1-Me-L-Trp	-41	-17	-33	8	-	-	b
TDO D-Trp	-33	-5	-17	9	16500	0	d
TDO-H55A							
L-Trp	-46	-16	-31	11	133	2.86	c
TDO-H55A							
1-Me-L-Trp	-48	-8	-34	15	59	0.048	b

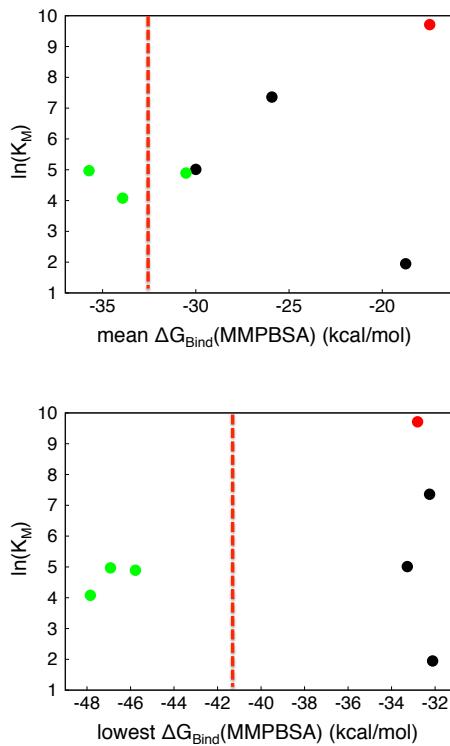


Figure 51 The mean (above) and lowest (below) binding free energies of each system as calculated by MM/PBSA plotted against each system's experimental K_M . Data for TDO with L-Trp, H55A TDO with L-Trp and H55A TDO with 1-Me-L-Trp are shown in green. TDO with D-Trp is shown in red. IDO with L-Trp, D-Trp and 1-Me-L-Trp are shown in black. The red lines indicate the calculated binding free energies for TDO with 1-Me-L-Trp, whose K_M could not be determined by experiment.

8.3.3. Substrate binding interactions

The active site hydrogen bonding network and substrate binding interactions of L-Trp in IDO were studied as part of the author's final year undergraduate research project.¹⁹⁰ The interactions identified in that study are shown in Figure 52 (top) and Figure 53 (top) in addition to the equivalent and additional interactions in TDO (bottom of both figures).

The hydrogen bonding interactions HB1, HB2, HB3 and HB4 are conserved across IDO and TDO, as are the salt bridges SB5 and SB6. HB7 and HB8 are unique to IDO and TDO respectively, although they are to some extent structurally analogous to each other. HB9, between the NE1 of the substrate and NE2 of His55 is unique to TDO with L/D-Trp and is of considerable interest due to the suggestion in earlier literature that proton transfer between these two nitrogens was a necessary step in catalysis.²⁰⁶

In the following analysis, the term ‘hydrogen bond population’ or ‘salt bridge population’ refers to the fraction of structures in which a particular interaction is considered present. The populations of each active site hydrogen bond and salt bridge interaction are shown in Figure 52 and Figure 53 for each simulation (each active site within each simulation in the case of TDO). The sum of all of the substrate binding interactions 4-9 (not including HB1,HB2 and HB3) for each simulation are plotted against the calculated binding free energies from that simulation in Figure 54. The IDO systems are plotted separately from the TDO systems because of the questionable validity of comparing free energies calculated with MM/PBSA for sufficiently different systems. In both IDO and TDO systems there is a clear correlation between the sum of the binding interaction populations and the calculated free energy. The higher the sum of the binding interaction populations, the lower the binding free energy. Most systems span approximately the same large range of values for the sum of binding interaction populations, however it is notable that maximum values for TDO with D-Trp and 1-Me-L-Trp are lower than for the other TDO systems.

Plotting only the sum of the salt bridge interaction populations against calculated binding free energies, a similarly strong correlation can be observed (Figure 55). Plotting interactions 4, 7, 8 and 9 (all hydrogen bonds with substrate) against calculated binding free energies also shows a significant (although slightly weaker) correlation of $R^2 = 0.40$ and 0.57 for IDO and TDO systems respectively (plot not reproduced here).

The wide spread of binding interaction populations over different production runs of the same system shows that these simulations have not converged with respect to active site structure. Most notably, some simulations form the salt bridges SB5 and SB6, which are maintained for almost the entire simulation, whereas in other simulations these salt bridges do not form at all. Insight into why this is can be gained by inspecting the crystal structures. The inhibitor bound to the IDO active site is not especially similar in shape to L-Trp and is uncharged so it would not be surprising if the active site conformation of the crystal structure is not optimal for substrate binding. In fact, in the crystal structure (with L-Trp bound in silico), the Arg231, which forms SB5 with the substrate, is pointing away from the active site out towards the solvent with a C-CZ distance (illustrated in Figure 53) of 6.9 Å. Similarly in active sites B and D of the TDO crystal structure, the equivalent residue (Arg117) is also pointing towards the solvent with a C-CZ distance of 5.6 Å. This is in contrast to active sites A and D in which the salt bridge SB5 is already formed with a C-CZ distance of 4.1 Å. In the simulations of TDO

with L-Trp, SB5 remains intact in active sites A and C throughout. In some simulations of TDO with L-Trp, SB5 forms at some point and then remains intact.

It seems that these 20ns simulations are not long enough to guarantee that the substrate and active site will settle into a low energy binding conformation. One of the simulations of IDO with L-Trp was extended to 50ns. Using structures from the 10th to 20th nanosecond of this simulation, the calculated binding free energy was -15 kcal/mol, the sum of the substrate binding interactions 4-7 was 1.43 and the sum of the salt bridge interaction populations alone was 0.95. When the structures from the 40th to 50th nanosecond were analysed, these values changed to -33 kcal/mol, 2.63 and 1.99 respectively, which are very similar values to those for the 20ns production run of IDO with L-Trp with the lowest calculated binding free energy. Although one cannot interpret too much from the extension of a single simulation, it seems likely that if all of these simulations were extended to the order of hundreds of nanoseconds much better convergence in terms of the active site structure and calculated binding free energies could be achieved.

Continuing with the assumption that the production runs with the lowest calculated binding free energies for each system are the most representative of the equilibrium ensemble of structures for that system, all further analysis presented here was performed on structures taken only from those simulations (and only from the appropriate active site in the case of TDO systems).

Table 18 shows the fraction of all structures (for each system) in which hydrogen bonds HB1, HB2 and HB3 are considered formed. For most systems, these three hydrogen bonds are formed for a significant fraction of each simulation. Simulations of IDO with O₂ but no substrate suggest that this hydrogen bonding network is present even in the absence of substrate. The most obvious outlier in this set is the simulation of TDO with D-Trp, in which HB3 is never formed and HB1 and HB2 are infrequently formed.

The data for the substrate binding interactions 4-9 are summarised in Table 19. Here D-Trp with TDO is again an outlier, as it never forms HB4 or SB6 in this simulation. When analysing all active sites of all simulations of D-Trp, SB6 is sometimes seen to form, but it never does so at the same time as SB5. It is also notable that 1-Me-L-Trp does not frequently form SB6 when bound to wild type TDO, but it does when bound to the H55A mutant.

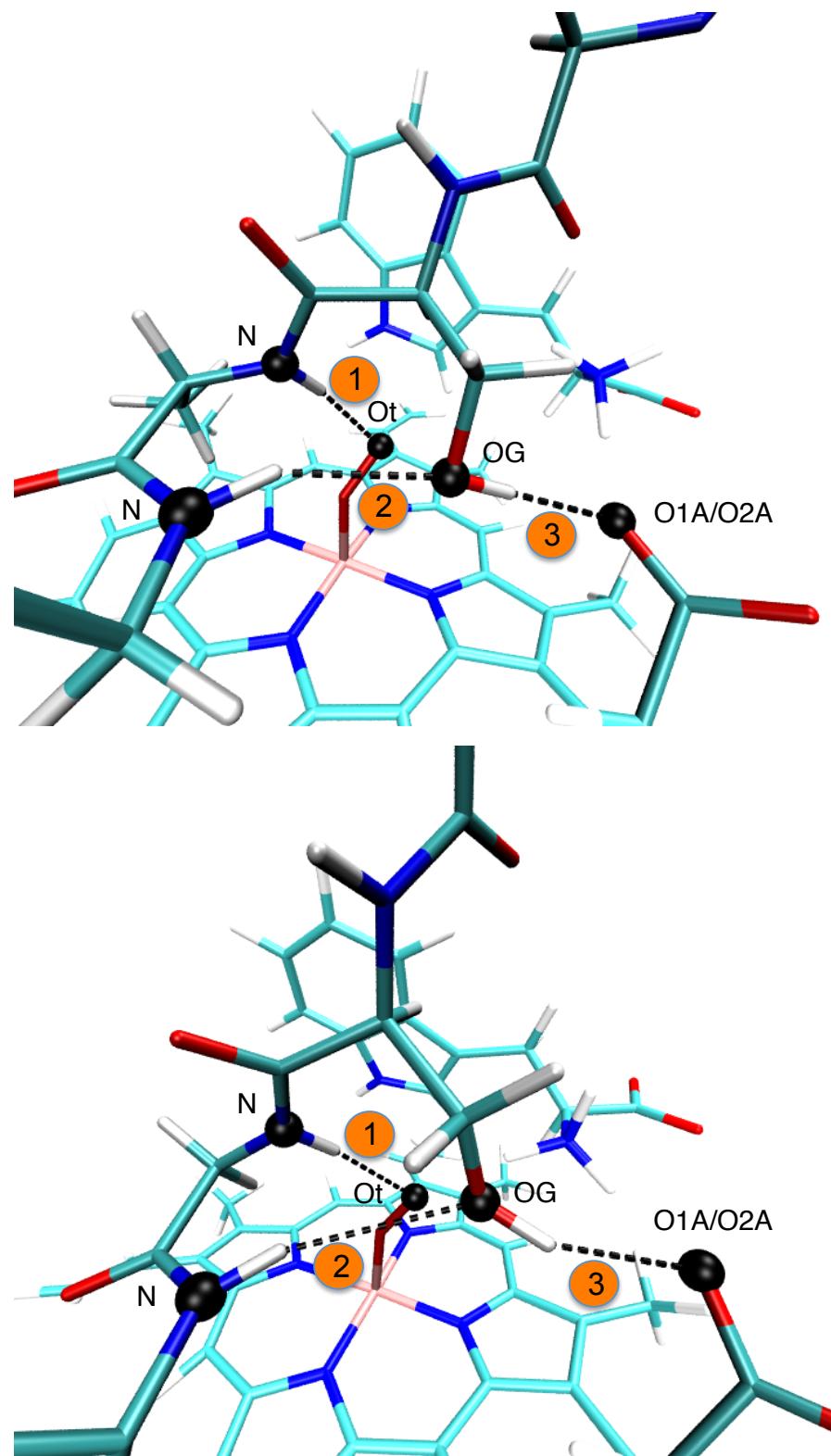


Figure 52 Conserved active site hydrogen bonding network for IDO (top) and TDO (bottom). Hydrogen bonds HB1, HB2 and HB3 are identified with orange circles.

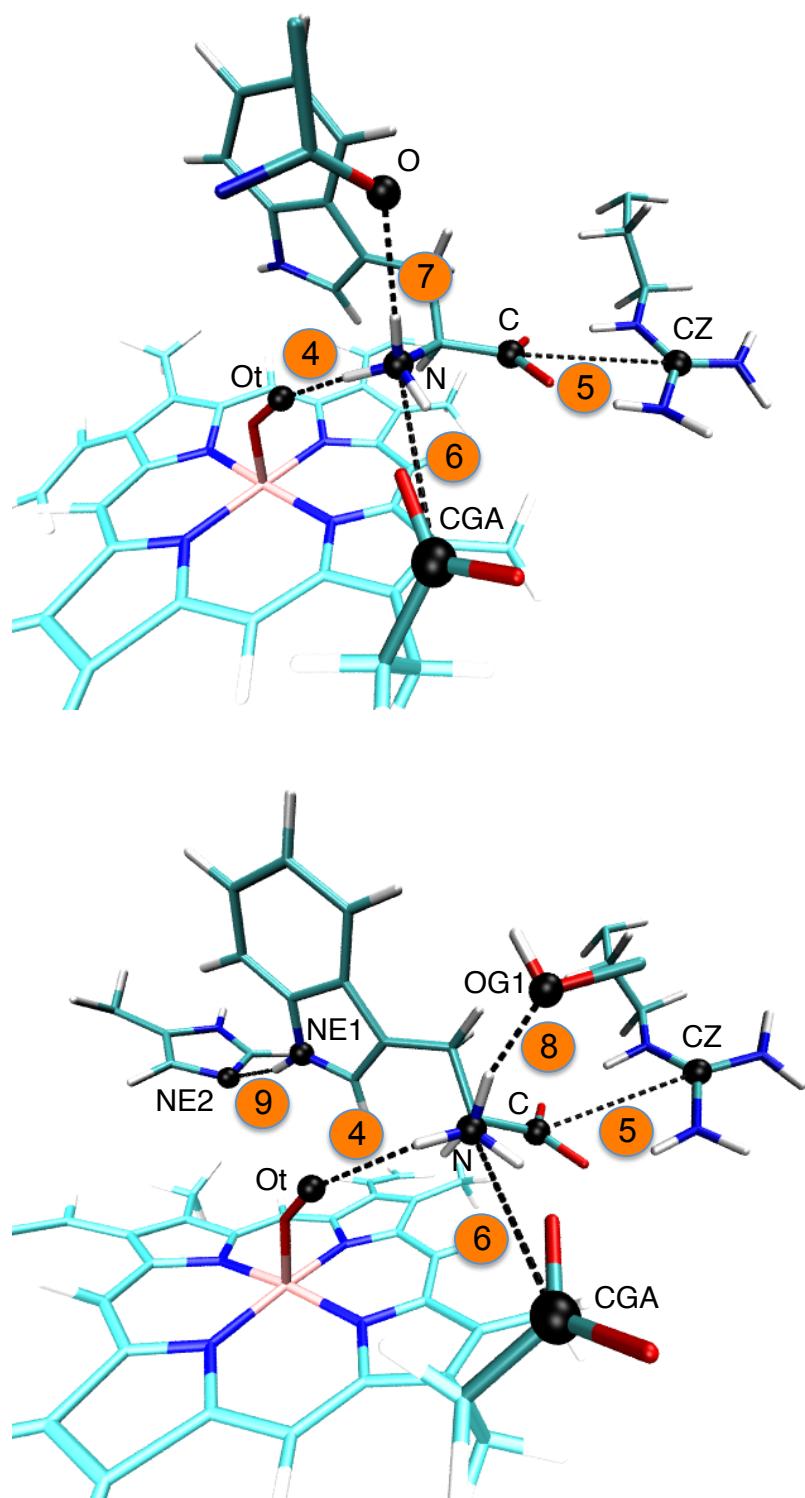


Figure 53 Substrate binding interactions in IDO (top) and TDO (below). Salt bridges SB5 and SB6 and hydrogen bonds HB4, HB7, HB8 and HB9 are identified orange circles.

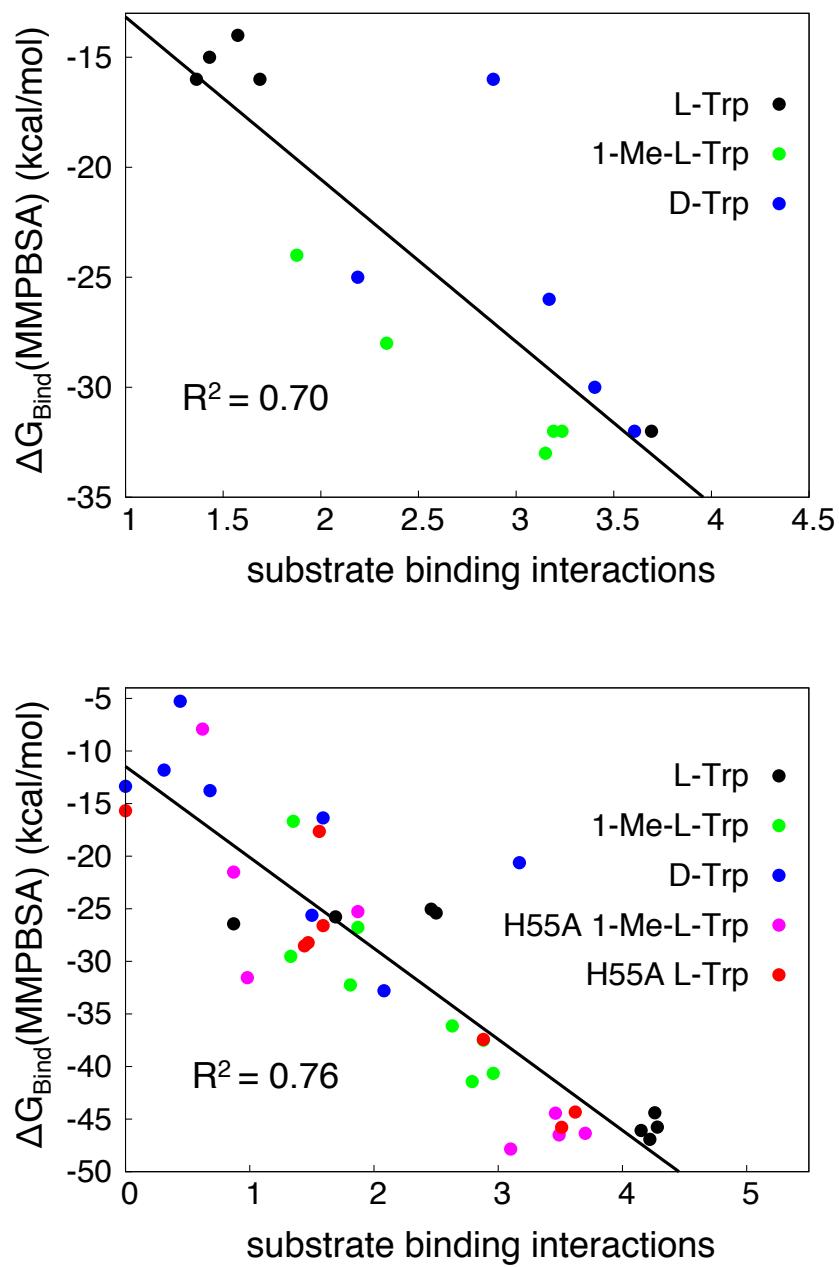


Figure 54 Sum of the populations of the substrate binding interactions 4-9 (as shown in Figure 53) for each production run of each system versus the binding free energy as calculated with MM/PBSA for that run. The population of a binding interaction is calculated as the fraction of the MD simulation in which the particular interaction is present. The data is separated into those for IDO (top) and TDO (bottom).

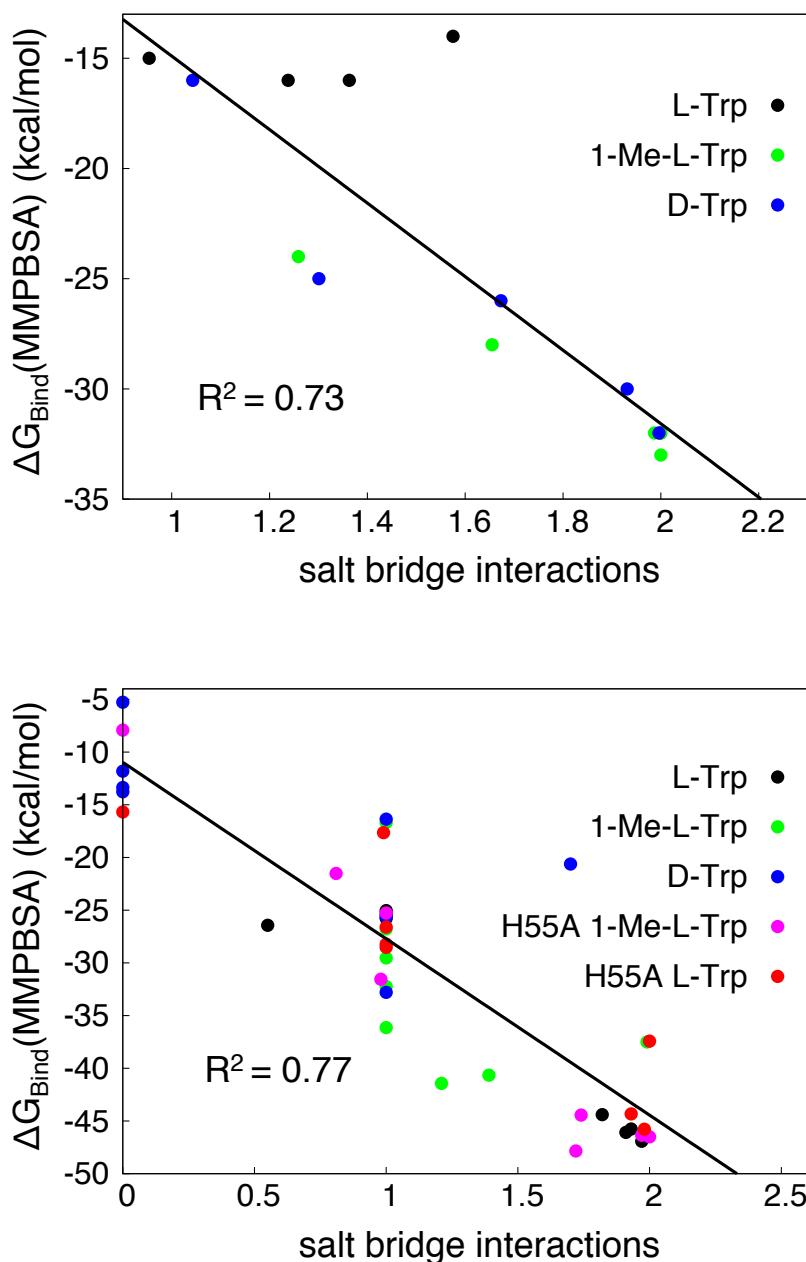


Figure 55 Sum of the populations of the salt bridge interactions SB5 and SB6 (as shown in Figure 53) for each production run of each system versus the binding free energy as calculated with MM/PBSA for that run. The population of a binding interaction is calculated as the fraction of the MD simulation in which the particular interaction is present. The data is separated into that for IDO (top) and TDO (bottom).

Table 18 Populations of active site hydrogen bond interactions shown in Figure 52. The structures used for this analysis were taken from the 10th to the 20th nanosecond of the simulation that gave the lowest calculated binding free energy for each system.

	HB1	HB2	HB3	sum 1-3
IDO Trp	0.95	0.77	0.88	2.60
IDO 1-Me-L-Trp	0.94	0.70	0.90	2.54
IDO D-Trp	0.73	0.51	0.85	2.10
IDO no substrate	0.70	0.56	0.85	2.12
TDO L-Trp	0.18	0.84	0.87	1.89
TDO 1-Me-Trp	0.54	0.12	0.95	1.61
TDO D-Trp	0.33	0.17	0.00	0.49
TDO H55A				
1-Me-Trp	0.78	0.49	0.95	2.22
TDO H55A				
L-Trp	0.42	0.80	0.85	2.06

Table 19 Populations of substrate binding interactions shown in Figure 53. The structures used for this analysis were taken from the 10th to the 20th nanosecond of the simulation that gave the lowest calculated binding free energy for each system.

	HB4	SB5	SB6	HB7	HB8	HB9	sum 4-9
IDO Trp	0.26	1.00	1.00	0.59	-	-	2.85
IDO 1-Me-L-Trp	0.35	1.00	1.00	0.23	-	-	2.58
IDO D-Trp	0.81	1.00	1.00	0.00	-	-	2.80
TDO L-Trp	0.76	1.00	0.97	-	0.82	0.67	4.22
TDO 1-Me-Trp	0.67	1.00	0.21	-	0.92	-	2.79
TDO D-Trp	0.00	1.00	0.00	-	0.52	0.56	2.08
TDO H55A							
1-Me-Trp	0.49	1.00	0.72	-	0.89	-	3.10
TDO H55A							
L-Trp	0.75	1.00	0.98	-	0.78	-	3.51

Evidence from the crystal structures and MD simulations shown here suggests that significant conformational change is associated with substrate binding and so the enzymes can be viewed as having open and closed forms. A long-standing debate in the study of substrate binding to proteins in general is whether substrate binding causes a particular conformational change (induced fit) or whether the substrate simply preferentially binds to and stabilises a pre-existing minority conformation (conformational selection).²⁰⁷ It is now suggested that induced fit and conformational

selection are two extremes of a spectrum of binding mechanisms.⁴⁴ As the conformational change described here involves the partial blocking of the active site, it is hard to see how the substrate could bind when the active site has already adopted the closed conformation. Consequently it seems likely that substrate binding goes by an induced fit mechanism at least to some extent. This is supported by the observation discussed above that in many of the MD simulations Arg231 (in IDO) and Arg117 (in TDO active sites B and D) start out exposed to the solvent but move in towards the active site to form a salt bridge with the substrate within several nanoseconds of simulation.

8.3.4. Substrate positioning for optimal reactivity

This section is best understood when read in conjunction with section 9.3.5. Energetic and structural analysis of the results of the QM/MM study of the heme dioxygenases showed a strong relationship between the starting structures used for adiabatic mapping and the resulting barrier heights. In particular, significant correlations were found between the barrier heights and weighted combination of distances ($Ot\text{-}C2$) + 0.33($Op\text{-}C2$) illustrated in

Figure 56. This parameter will be referred to as λ .

For each system, the trajectory of the MD simulation with the lowest calculated binding free energy was analysed with respect to λ , in an attempt to find a relationship between substrate positioning in the active site and reactivity. Histograms of the distributions of λ for all systems are shown in Figure 57. The distributions can be divided into 3 groups: those which lie furthest to the left (shown in black), those furthest to the right (TDO with 1-Me-L-Trp in blue and TDO with D-Trp in purple) and the one in between the two extremes (IDO with 1-Me-L-Trp).

The line of best fit (from linear regression) that ran through the points on the graph plotting the QM/MM barrier heights for all adiabatic mapping pathways against λ had the formula $\Delta E = 24.7\lambda + 75$. According to the Eyring equation, a rate constant of 0.01s^{-1} would correspond to a barrier height of 20 kcal/mol at 298K. Reading from the line of best fit, for a pathway to have a barrier height of less than 20 kcal/mol, the starting structure for adiabatic mapping would have to have a value of λ lower than 3.87 Å. The line $x = 3.78$ is shown in Figure 57.

Table 20 shows the proportion of each simulation that has a value of value of λ lower than 3.87 Å. The two systems in which λ is never or almost never lower than 3.87 Å are also the two systems which were found experimentally not to turn over. IDO with 1-Me-L-Trp, which has the lowest non zero rate constant of all the systems shown here also has a fairly low proportion of structures in which λ is lower than 3.87 Å. The H55A mutant with L-Trp has a fairly low rate constant and so according to this theory of optimal substrate positioning would be expected to have a slightly lower proportion of structures in which λ is lower than 3.87 Å. This slight anomaly notwithstanding, there appears to be a significant relationship between the positioning of the substrate in the active site as observed in MD simulations and experimentally observed rate constants.

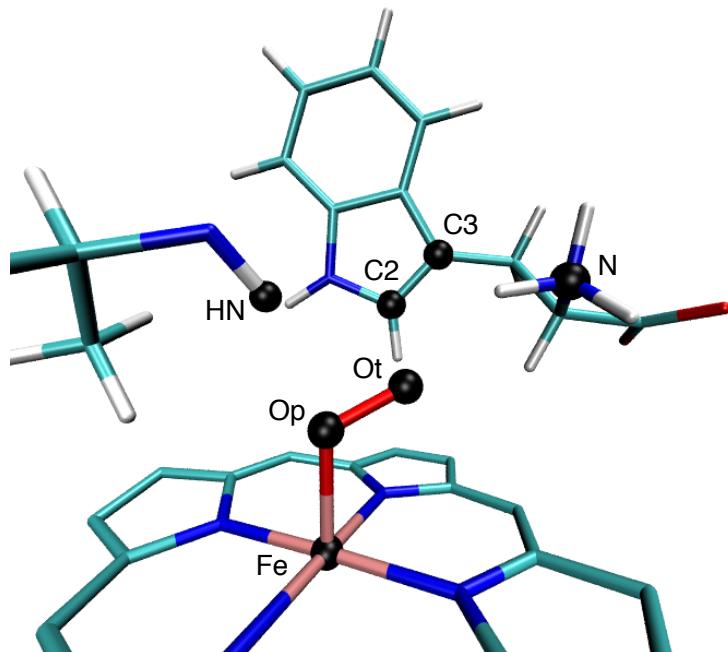


Figure 56 Key atoms used in the analysis of substrate positioning. In particular, the proportion of structures with low values of the parameter $\lambda = (Oe-C2) + 0.33(Ot-C2)$ was found to be related to barrier heights.

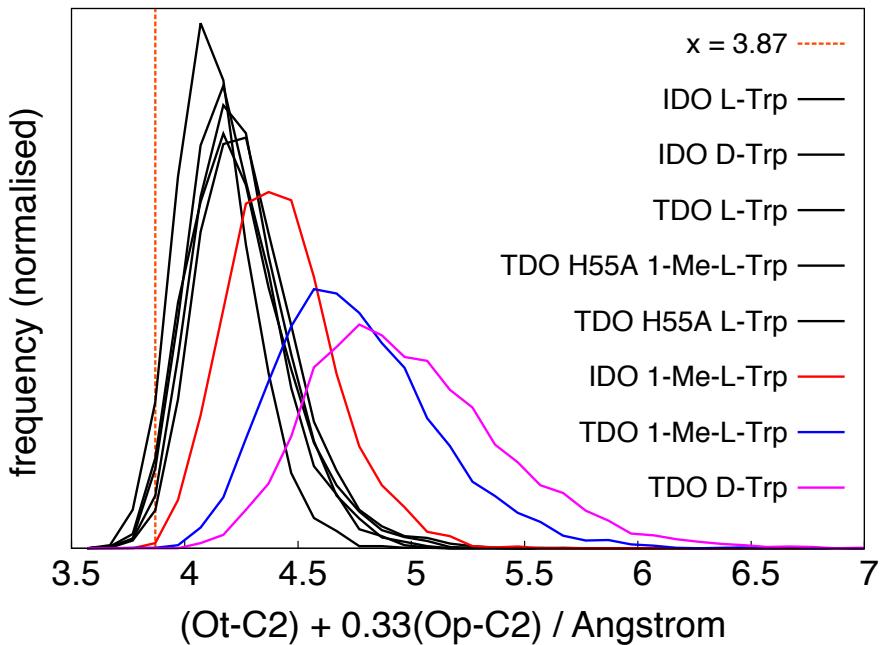


Figure 57 Histograms showing the distribution of values of the parameter $(Ot\text{-}C2) + 0.33(Op\text{-}C2)$ for each system. The structures used for this analysis were taken from the 10th to the 20th nanosecond of the simulation that gave the lowest calculated binding free energy for each system. Analysis of structures and barrier heights from QM/MM simulations suggest that a value of $(Ot\text{-}C2) + 0.33(Op\text{-}C2)$ less than 3.87 Å is required for the overall barrier to be less than 20 kcal/mol.

Table 20 Proportion of structures from MD simulations with values of λ lower than 3.87 Å compared to the experimental rate constants. $\lambda = (Ot\text{-}C2) + 0.33(Ot\text{-}C2)$.

	$(Ot\text{-}C2) + 0.33(Ot\text{-}C2)$ % < 3.87 Å	exp. kcat (s-1)
IDO L-Trp	2.14	1.4
IDO 1-Me-L-Trp	0.24	0.027
IDO D-Trp	4.20	3.93
TDO L-Trp	2.94	19.5
TDO 1-Me-L-Trp	0.04	-
TDO D-Trp	0.00	-
TDO H55A 1-Me-L-Trp	8.70	2.86
TDO H55A L-Trp	4.92	0.048

Figure 58 shows average structures from MD simulations of TDO with L-Trp and with 1-Me-L-Trp superimposed on the crystal structure (active site A). The average structure for the L-Trp simulation is very similar to the crystal structure, which reflects well on the quality of the force field used. The average positioning of 1-Me-L-Trp is slightly different. The C2 carbon has moved away from the terminal oxygen (consistent with the larger observed values of λ). His55 has maintained its position in the active site, so it appears that the shifting of the 1-Me-L-Trp could be largely due to steric clash with between the methyl group and the histidine.

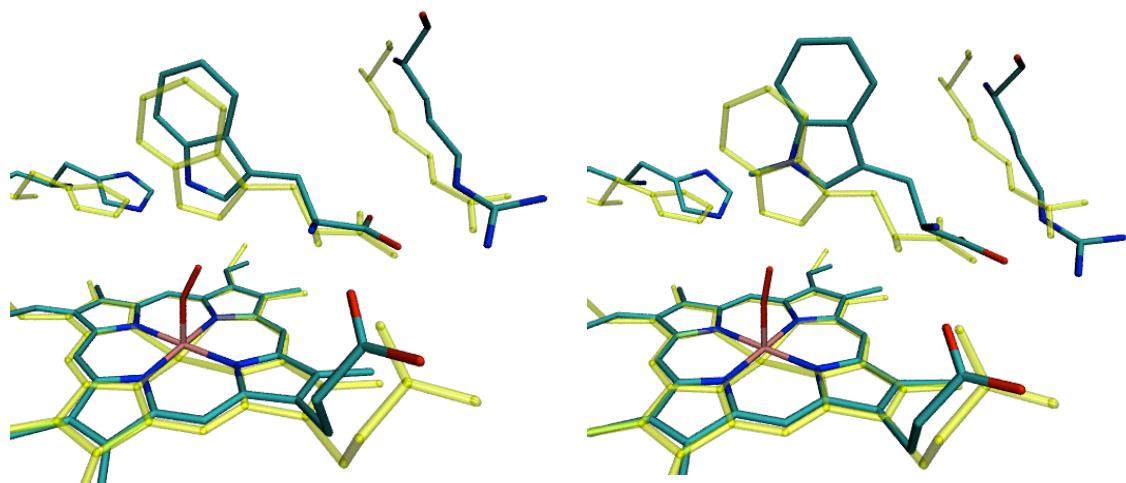


Figure 58 Average structures from single MD production runs of TDO with L-Trp (left) and TDO with 1-Me-L-Trp (right) superimposed onto the crystal structure 2NW8 (yellow).

Chapter 9. QM/MM study of reactivity in IDO and TDO

9.1. Introduction

Recent computational and experimental work has altered the consensus on the mechanism of the heme dioxygenases.^{168,177} The current prevailing mechanism was outlined in an computational paper on TDO by the group of Morokuma and is believed to be credible on the basis of good agreement between calculation and experiment (as discussed in section 7.2).¹⁷⁷ This mechanism involves radical addition of the terminal oxygen of ferric heme bound superoxide to the C2 carbon of the substrate to form an alkylperoxo intermediate, followed by homolytic cleavage of the O-O bond to form compound II and an epoxide species. The first and second steps are proposed to occur in the singlet and triplet states respectively. The second step was calculated to be cleanly rate limiting, so further steps will be neglected in present study.

The main aim of the work described in this chapter is to gain insight into the variation in substrate specificity between IDO and TDO in terms of differences in barriers to reaction. To this end, reaction profiles were calculated for IDO and TDO with L-Trp, D-Trp and 1-Me-L-Trp using QM/MM adiabatic mapping. In addition, answers to further questions about the mechanism are sought, such as: how and where does spin state occur during the reaction, whether the alkylperoxo species is an intermediate (as calculated by Morokuma) or a transition state (as calculated in another QM/MM study²⁰⁸), can further evidence against alternative reaction mechanisms be found and what role the enzyme plays in catalysis.

9.2. Computational details

9.2.1. Selecting structures for QM/MM calculations

Adiabatic mapping with DFT QM/MM is a computationally expensive tool, so it is generally only feasible to calculate of the order of tens of pathways. The barrier heights

obtained for individual pathways tend to be highly sensitive to variations in the geometry of the starting structures.^{50,52} As only the lowest energy pathways will contribute significantly to the ensemble averaged barrier heights, it is desirable to select starting structures from MD simulations that are likely to lead to low energy pathways, in order to improve convergence of results. Previous work has shown that strong correlations can exist between the values of simple geometric parameters of the starting structures and the barrier heights of individual pathways.^{50,52} As such, a set of criteria can be devised in order to screen MD trajectories for so-called pre-reactive structures. Provided that a significant proportion of all structures from MD trajectories satisfy these criteria, no significant bias is introduced into the calculated barrier heights by this screening process as the selected structures will all represent thermally accessible states of the system.

In this study, two geometric parameters were chosen in order to screen the MD trajectories for pre-reactive structures. The first was the distance $R(C2-Ot)$ between the terminal oxygen and the C2 carbon of the substrate. This was an obvious choice given that it corresponds to the bond formed in the first step of the putative reaction mechanism investigated here. The second parameter was the distance $R(Ot-N)$ between the terminal oxygen and the NH_3^+ nitrogen of the substrate. (See Figure 70 for an illustration of both of these distances). This was chosen because evidence from ENDOR spectroscopy and analysis of the MD trajectories described in section 8.3.3 suggest that, in the Michaelis complex, a hydrogen bond is present between the substrate NH_3^+ group and the terminal oxygen (Ot).¹⁶⁵ For a structure from MD to be eligible for selection both $R(C2-Ot)$ and $R(Ot-N)$ distances had to be less than 3 Å. From this subset of the MD trajectory snapshots, structures were randomly chosen as starting points for adiabatic mapping.

It is worth noting briefly here that the value of $R(C2-Ot)$ of the starting structures correlated well with the calculated barrier heights, whereas this was not the case for the value of $R(Ot-N)$. As such, the screening criteria used here has proven to be sub-optimal, as will be discussed in more detail in 9.3.5.

9.2.2. System truncation and partitioning

Once a suitable structure was chosen from MD simulations, the total system was truncated in order to comply with the maximum number of atoms permitted in Tinker. This had the added benefit of reducing the computational expense due to the MM point

charges included in the QM part of the QM/MM calculations. No protein atoms were removed during this truncation. For IDO, all ions and whole water molecules further than 4 Å from the protein were removed. For TDO all ions and whole water molecules further than 20 Å from the heme in active site A were removed. Because most of the ions were removed during this process the systems were no longer neutral. As the ions are generally located in the solvent or on the surface of the protein and not particularly near the active site, it was assumed that this will not significantly affect the results.

The QM region for all systems consisted of the full substrate, O₂, a truncated heme (without propionate or methyl groups) and imidazole (see Figure 66). Larger QM regions were not investigated. However similar heme truncation schemes have been used and found to be acceptable in a number of other QM and QM/MM studies.^{177,209–211} For calculations on TDO, only the relevant atoms in active site A were included in the QM region. The link atom method was used to treat the boundaries between the QM and MM regions and the charges of MM atoms near the boundary region were adjusted as described in section 1.1.3.

The active region (in which atoms are allowed to move during optimisations) was defined as containing all atoms belonging to residues with at least one atom within 6 Å of the QM region. The frozen region contained all other atoms. The QM, active and frozen regions are illustrated in Figure 59.

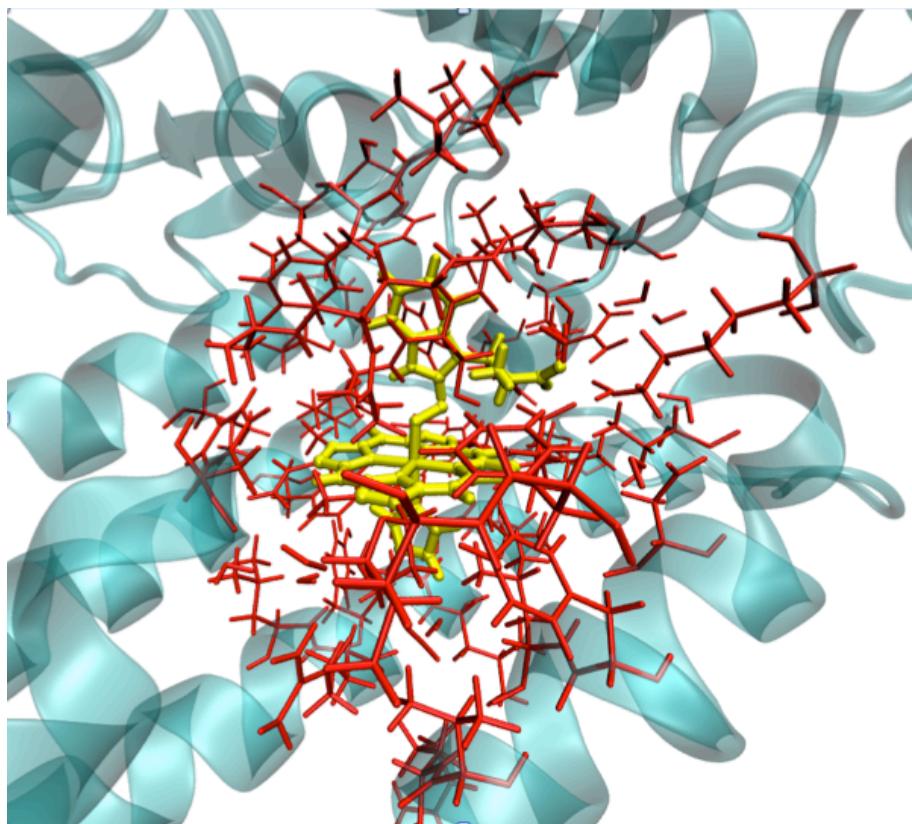


Figure 59 Partitioning of the QM/MM system. The QM and active regions are shown in yellow and red respectively. All other atoms are in the frozen region. For clarity, the atoms of the frozen region are shown by a cartoon representation of their secondary structure.

9.2.3. QM/MM details

All QM/MM calculations were performed with QoMMMa (our in-house QM/MM code) which called Jaguar version 7.8 for all QM calculations and Tinker Version 6.1 for all MM calculations.^{113,138,139} B3LYP with Grimme's empirical dispersion correction version 2 (as implemented directly in QoMMMa) was used for all QM/MM calculations.^{97,98} Broken-symmetry unrestricted DFT was used for all open shell calculations. To ensure convergence to the desired open shell electronic states, extensive use of Jaguar's 2spin, fragment and istate commands were used. The basis set used for all optimisations was LACVP*, which consists of the 6-31G basis set with polarisation functions on all non hydrogen atoms for atoms H-Ar and the Hay and Wadt effective core basis set for heavier atoms (Fe in this case).^{101,212} The triple zeta contraction of LACVP, with diffuse functions on all atoms and polarisation functions on all non transition metal atoms, LACV3P++** was used for all single point calculations. Unless otherwise stated,

all energies given in this chapter are from single point calculations with the LACV3P++** basis set performed on structures optimised with the LACVP* basis set. As with the MD simulations, the CHARMM27 force field was used for the MM region.¹⁹¹

9.2.4. Minimizing the Michaelis complex

The truncated starting structures were initially minimized with a MM force field in NAMD.¹⁹³ Some harmonic restraints were imposed on these minimisations to ensure that the MM starting structure sufficiently resembled the minimum along the QM/MM reaction pathway for the first step of the reaction. This was necessary due to deficiencies in the force field such as a lack of Fe-Op-Ot angle and N*-Fe-Op-Ot dihedral terms (where N* represents any of the Fe coordinating nitrogens). The values of the NA-Fe-Op-Ot dihedral angle, the Fe-Op-Ot angle and the Ot-C2 distance were restrained to 45°, 120° and 3 Å respectively.

QM/MM optimisations with no restraints (other than the freezing of atoms in the frozen region) were then performed using restricted open shell DFT (RODFT) in the triplet state. With the QM regions extracted from these structures, Jaguar was used on its own to generate initial guess broken-symmetry unrestricted wavefunctions for the various different electronic configurations to be investigated. The 2spin option of the atomic section was used to constrain an excess of 1 alpha spin on Fe and an excess of 1 beta spin on O₂. The istate option of the gen section was used to explore different low-lying d-orbital configurations. The resulting orbitals were used as initial guesses for QM/MM optimisations of the various states, starting from the RODFT triplet optimised structures. Closed shell singlets were calculated using restricted DFT. Full QM/MM optimisations of various different states were only performed for a small number of IDO with L-Trp structures. The character and energy ordering of the states was consistent for all structures calculated, so it was not deemed necessary to perform these calculations for all structures to be taken on to adiabatic mapping.

9.2.5. Adiabatic mapping

Starting with the RODFT triplet optimised starting structures and the initial guess wavefunction for the singlet ground state, QM/MM adiabatic mapping was performed on each structure. A harmonic restraint of 2000 kcal/mol/Å² was applied to the reaction coordinate R(Ot-C2). The minimum of this restraint was set to 3.0 Å and the active region was fully optimised. The value of this minimum was decreased in steps of 0.1 Å until a value of 1.4 Å had been reached. As is common with adiabatic mapping, many of the paths exhibited some discontinuities in their surfaces, usually due to some seemingly trivial movement of a part of the MM region. To smooth these out, adiabatic mapping was performed in the backward direction starting from the structures optimised at R(Ot-C2) values of 2.0 Å from adiabatic mapping in the forward direction. The reported profiles and energies are from combinations of the backward mapping from 2.0 Å to 3.0 Å and the forward mapping from 2.0 to 1.4 Å.

For several IDO with L-Trp pathways, the same procedure was followed starting in the triplet state with the d_{yz} orbital singly occupied (the triplet state equivalent to the singlet ground state). As the singlet pathway was significantly lower in energy for each pathway calculated, it was not considered necessary to calculate the triplet pathway in every case.

The structures representing the product minimum along the R(Ot-C2) reaction coordinate (which were the structures restrained to R(Ot-C2) = 1.5 in all cases), were used as starting structures for the second step of the reaction. For this step, a harmonic restraint was imposed on the distance R(Ot-Op). The same procedure was followed as with the first step of the reaction with the minimum value of the harmonic restraint being increased in steps of 0.2 from 1.45 to 2.65 Å. As with the first step, several IDO with L-Trp pathways were calculated in the triplet and singlet states, however as the triplet pathway was much lower in energy for each pathway calculated, only the triplet pathway was calculated for the second step in every case.

Single point calculations with the LACV3P++** basis set were performed on the five stationary points found along the pathways. These stationary points were taken to be the lowest and highest point from the relevant sections of the pathways. For one test pathway, the three minima were fully reoptimised without the reaction coordinate restraints. The difference in energy between the lowest point of the relevant region and the structure optimised without restraints was less than 0.05 kcal/mol in each case,

indicating that the lowest energy structures along the adiabatic mapping pathways are good approximations of the true minima. No such test was performed for the two maxima along each pathway. From visual inspection of the pathways it seems likely that a much larger error is introduced by not fully optimising the transition states. This is because the curvature of the surface is much higher in the regions of the maxima than in those of the minima.

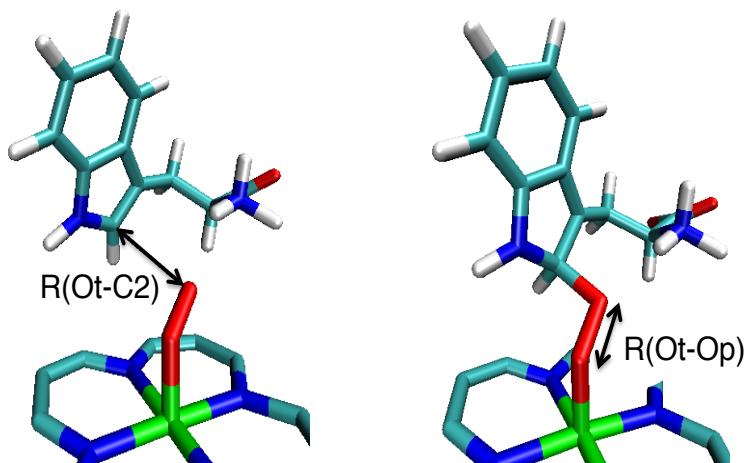


Figure 60 The reaction coordinates used to drive the adiabatic mapping of step I (left) and step II (right) of the heme dioxygenase reactions.

9.3. Results and discussion

9.3.1. Electronic structure of the Michaelis complex

The electronic structure of oxy-heme complexes has been extensively debated since 1936 when Linus Pauling first tried to rationalise the changes in observed magnetic properties of hemoglobin that accompany binding of O₂ (deoxy-hemoglobin is paramagnetic and oxy-hemoglobin is diamagnetic).²¹³ Pauling suggested a valence model in which the ground state is a singlet ferrous oxide Fe(II)-O₂, with no unpaired electrons (closed shell singlet). Another model, proposed by Weiss, describes a singlet ferric superoxide Fe(III)-O₂^{•-} species with one unpaired electron in an Fe(III) d-orbital and the other on the superoxide with opposite spin (with antiferromagnetic coupling accounting for the experimentally observed diamagnetism).²¹⁴ Several other models and refinements have been proposed over the years.^{215,216}

Previous QM and QM/MM DFT studies have provided support for the Weiss ferric superoxide model.²¹⁷ However, multireference and muticonfigurational *ab initio* studies have shown that oxy-heme systems cannot be fully described by any single valence model, rather the system has significant multiconfigurational character.^{218,219} It is beyond the scope of this work to enter into the pursuit of the most accurate description possible of the electronic structure of oxy-heme. For practical reasons we have chosen to limit ourselves to the use of DFT. We have been careful to ensure that we have modelled the correct states within this framework while bearing in mind the known limitations of DFT in describing oxy-heme systems.

Heme systems, like many transition metal systems, have several low lying near degenerate states.²²⁰ For systems with unpaired electrons, these states often differ in which metal d-orbitals the unpaired electrons occupy and in the spin orientations of those electrons. Figure 61 shows the various Fe d-orbitals in the present system as well as the singly occupied O₂ π* orbital. Various configurations of singly occupied d-orbitals were explored in search of the ground state and other low-lying states. The results are illustrated in Figure 62.

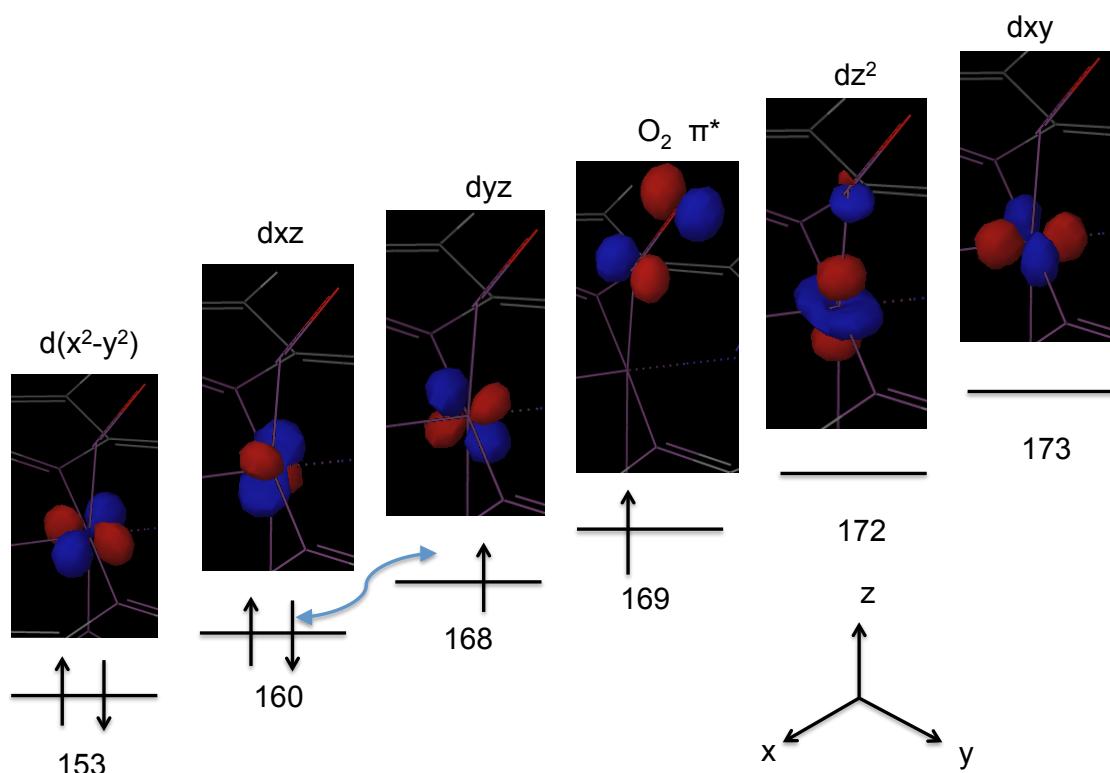


Figure 61 Energy ordering and occupation of the Fe d-orbitals. The singly occupied O₂ π* orbital is also shown. The low lying singlet and triplet states differ in the which of the d_{xz} or d_{yz} orbitals are singly occupied and in the relative spin orientation of the singly

occupied orbitals. The orbitals shown here are from a triplet RODFT calculation on the Michaelis complex.

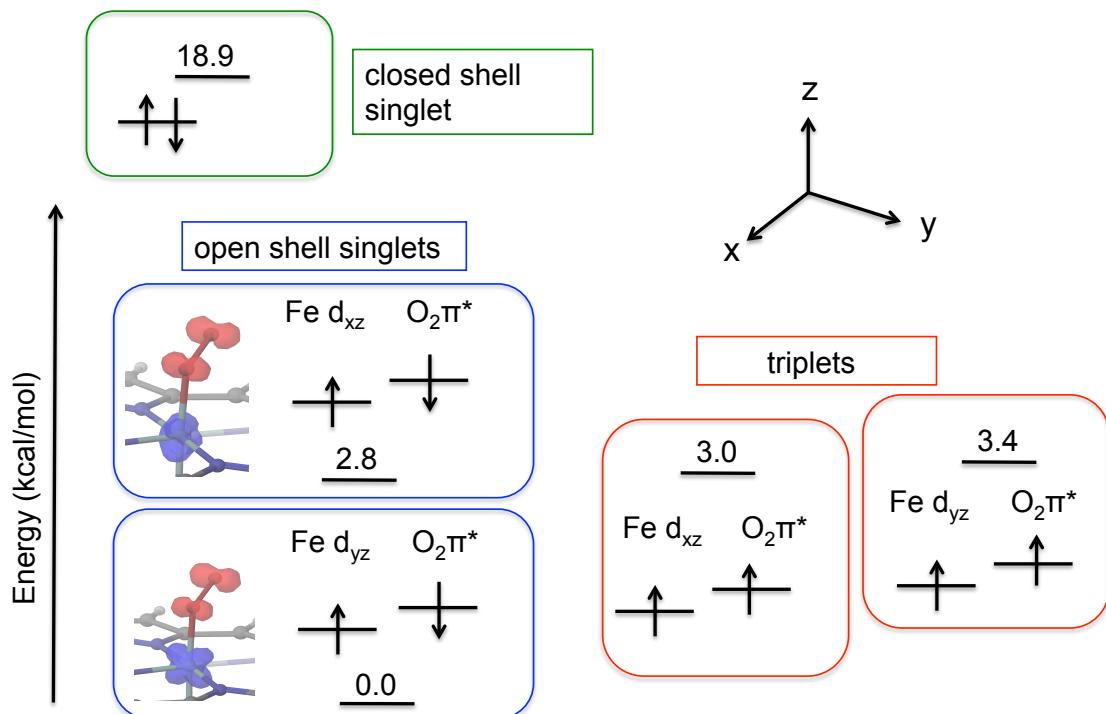


Figure 62 Open shell singlet ground state and low lying excited singlet and triplet states. Excess alpha and beta spin densities are depicted as blue and red isosurfaces respectively. Spin densities calculated with Mulliken population analysis. Each state shown above was optimised separately from a single starting structure taken from an MD simulation of IDO with L-Trp.

The ground state for the Michaelis complex was found to be a singlet with two unpaired electrons with opposite spin as predicted by the Weiss mode. The configuration found here is consistent with the ground state found in a QM/MM study of oxy-myoglobin.²¹⁷ Another singlet was found at around 2 kcal/mol above the ground state. These states differ in the identity of the singly occupied Fe d-orbital. The corresponding triplet states are energetically degenerate, lying around 3 kcal/mol above the ground state. The closed shell singlet state (corresponding to the Pauling model), calculated with restricted DFT was found to lie 19 kcal/mol higher in energy than the ground state. If unrestricted DFT was used without providing any initial guess, the wavefunction converged to the closed shell result with an $\langle S^2 \rangle$ value of 0, highlighting the care that needs to be taken with these systems to converge to the desired state.

The correct $\langle S^2 \rangle$ for an open shell singlet, as calculated using Equation 17, is 1. The singlet states suffer from spin contamination, with $\langle S^2 \rangle$ values of 0.95 and 1.08 for the ground and first excited states respectively. This could be corrected with the use of spin projection, although we have not done this. We do not expect this to introduce large errors as the effect of spin contamination on the relative energies of the spin states was found to be slight in the QM/MM study of oxy-myoglobin previously mentioned.²¹⁷

$$\langle S^2 \rangle = S_z(S_z + 1) + N_\beta - \sum_{ij}^{N_{MO}} \langle \phi_i^\alpha | \phi_j^\beta \rangle^2$$

Equation 17 The spin expectation value, $\langle S^2 \rangle$, of an unrestricted wavefunction. S_z is the number of unpaired electrons divided by two. N_β is the number of electrons with beta spin. ϕ_i^α and ϕ_j^β are the orbitals occupied by electrons with alpha and beta spin respectively.⁸

9.3.2. Reaction mechanism

Our calculations support the mechanism outlined in the ONIOM study of TDO undertaken in the group of Morokuma.¹⁷⁷ This mechanism involves radical addition of the terminal oxygen in the ferric superoxide state to the C2 carbon to form a alkylperoxo intermediate followed by homolytic cleavage of the O-O bond to form an epoxide and compound II. Alternative mechanisms were investigated in this work and are discussed in section 9.3.8. The ONIOM study found all further steps along the reaction pathway to be significantly lower in energy than the transition states for these first two steps. As it seemed clear that these first two steps are rate limiting, we did not model any of the further steps in this study.

Figure 63 shows representative reaction profiles for the first two steps along the singlet (blue) and triplet (red) surfaces of IDO with L-Trp. Table 21 summarises the changes in distribution of excess spin density over the course of the reaction. This is also illustrated graphically by the isosurfaces shown in Figure 66. As discussed in the preceding section, the ground state of the Michaelis complex is an antiferromagnetically coupled open shell singlet. In this state, there is an excess of alpha spin on the iron and an excess of beta spin on the dioxygen (slightly weighted towards the terminal oxygen).

The first transition state for the singlet pathway was found to be lower than that for the triplet pathway (relative to the singlet Michaelis complex) for each of the 3 pathways

in which both states were modelled. It was concluded that the first step most likely proceeds via the singlet pathway. The triplet pathway for the first step is however not very much higher in energy, so the reaction may sometimes proceed by this pathway, spin crossover having occurred at some point in the Michaelis complex.

On forming the alkylperoxo intermediate, the singlet and triplet states become degenerate. The excess alpha spin density is centred on the iron, whereas the excess beta spin density is centred on the C3 carbon. Spin crossover at this point in the reaction is discussed in section 9.3.3.

The second transition state for the triplet pathway was found to be significantly lower than that for the singlet pathway for each of the 3 pathways in which both states were modelled, so we conclude that this step will almost certainly proceed in the triplet state. Finally, an epoxide intermediate and compound II are formed. This formation of compound II is in agreement with the experimental detection of a ferryl species during IDO turnover.¹⁷⁴

Figure 65 shows a schematic reaction profile for these first two steps of the reaction with energies given for a representative pathway of IDO with L-Trp.

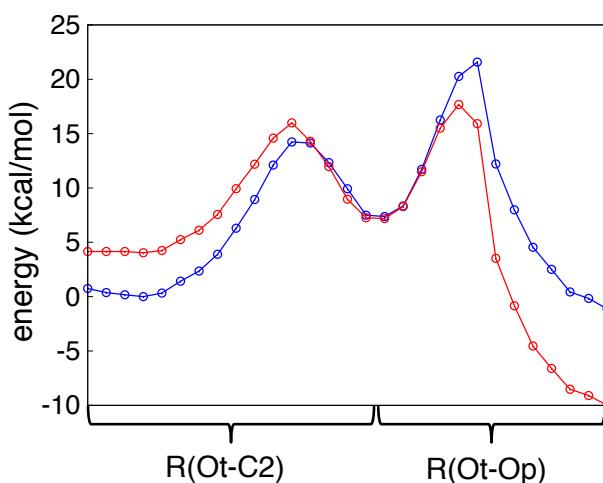


Figure 63 Typical potential energy profiles from adiabatic mapping of IDO with L-Trp on the singlet (blue) and triplet (red) surfaces. Note that the x-axis represents step number, not actual values of the reaction coordinate.

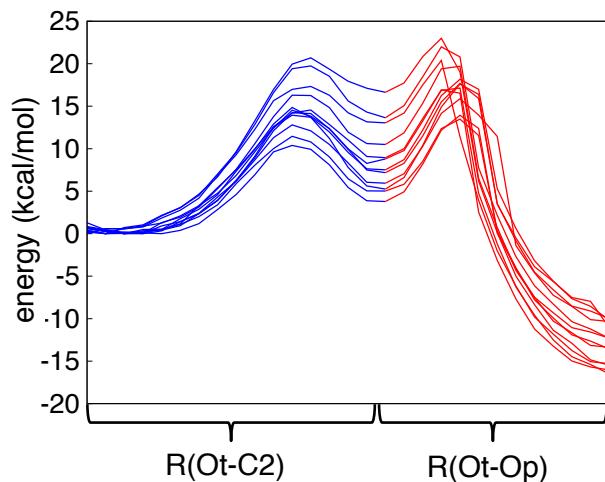


Figure 64 Potential energy profiles from adiabatic mapping of IDO with L-Trp on the singlet (blue) and triplet (red) surfaces. Different profiles were generated starting from different structures taken from MD simulations.

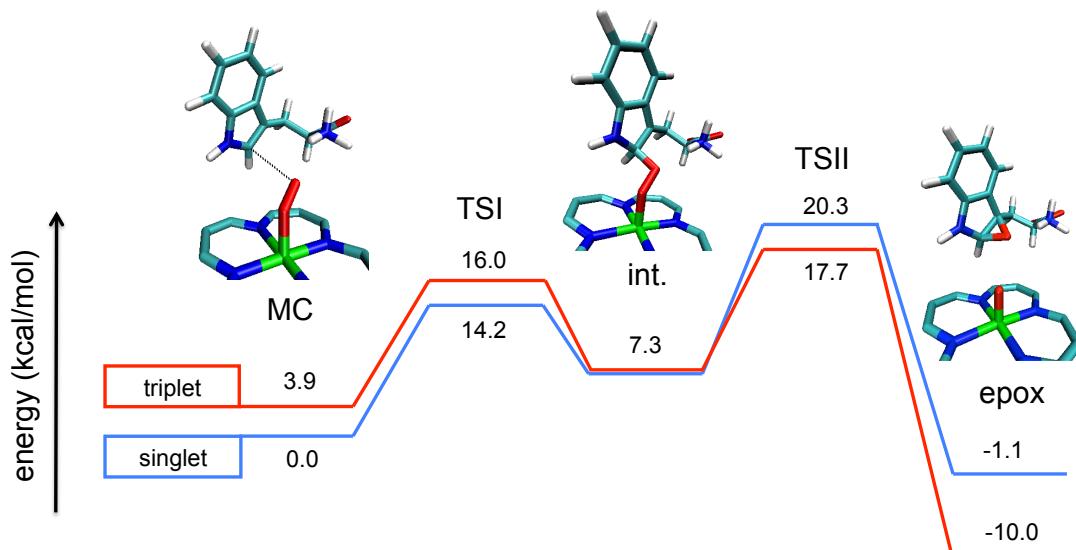


Figure 65. Typical schematic potential energy profiles from adiabatic mapping of IDO with L-Trp on the singlet (blue) and triplet (red) surfaces.

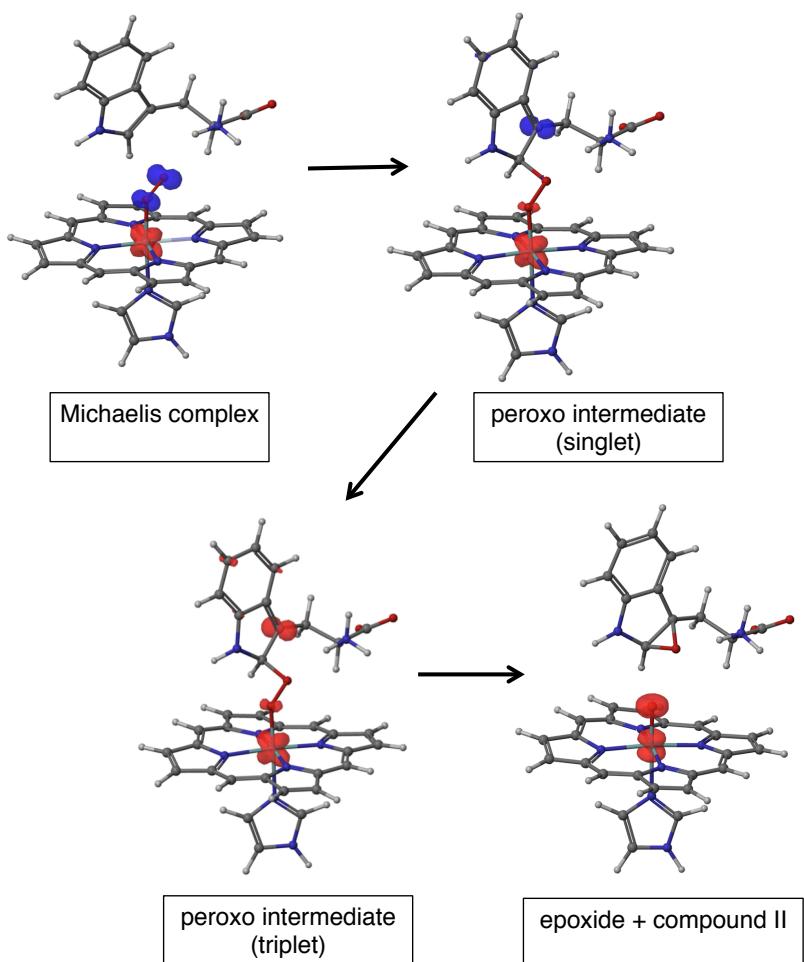


Figure 66 Isosurfaces representing densities of excess alpha spin (red) and excess beta spin (blue) at the three minima along one of the QM/MM adiabatic mapping pathway of IDO with L-Trp. Spin densities calculated with Mulliken population analysis.

Table 21 Spin densities from Mulliken population analysis at the stationary points along one of the QM/MM adiabatic mapping pathways of IDO with L-Trp. Values given only for atoms which have large spin populations at some point along the reaction path.

	Fe	Op	Ot	C3
Michaelis (singlet)	1.07	-0.40	-0.61	-0.02
TSI (singlet)	1.01	-0.10	-0.34	-0.37
intermediate (triplet)	1.01	0.08	0.02	0.59
TSII (triplet)	1.04	0.57	-0.23	0.40
CpII + epoxide (triplet)	1.21	0.86	0.00	0.00

9.3.3. Spin Crossover at the alkylperoxo intermediate

Given that the Michaelis complex is a singlet but compound II is a triplet, the system must change spin state at some point during the reaction. The most likely place for this spin state change to occur is at the lowest energy point at which the singlet and triplet surfaces cross, known as the minimum energy crossing point (MECP).²²¹ Figure 67 illustrates the intersection of a pair of hypothetical singlet and triplet surfaces.

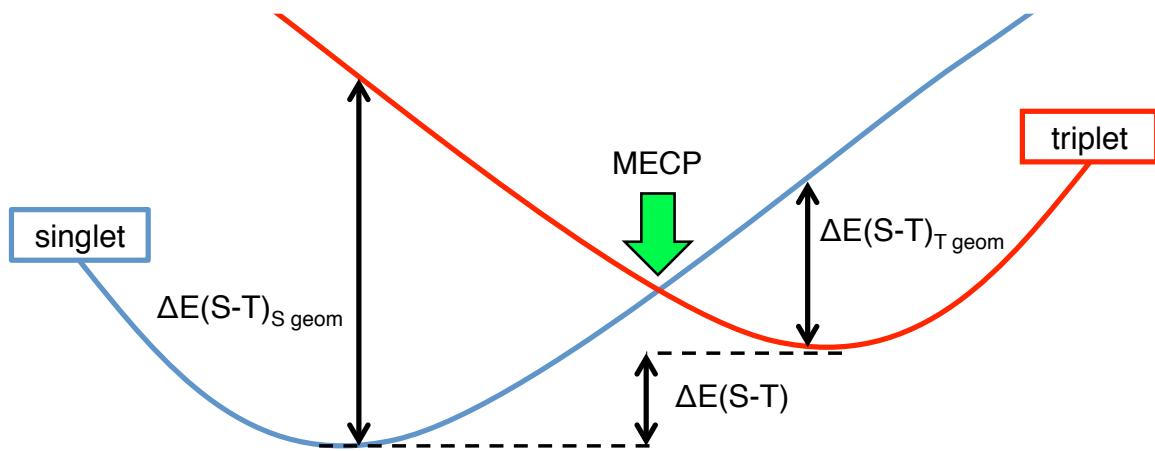


Figure 67 Diagram illustrating the crossing of two hypothetical singlet and triplet surfaces. $\Delta E(S-T)_{S\text{ geom}}$ is the difference in energy between the singlet and triplet states at the optimised singlet geometry, $\Delta E(S-T)_{T\text{ geom}}$ is the difference in energy between the singlet and triplet states at the optimised triplet geometry, $\Delta E(S-T)$ is the difference in energy between the singlet and triplet states at their respective optimised geometries. MECP is the minimum energy crossing point.

Looking at Figure 63 it appears that the MECP for this system is in the region of the alkylperoxo intermediate. In order to investigate the overlap of singlet and triplet surfaces in the vicinity of the alkylperoxo intermediate, the relative energies between the two surfaces at various structures were calculated (structures taken from one adiabatic mapping pathway only). These were: the difference in energy between the singlet and triplet states at the optimised singlet geometry ($\Delta E(S-T)_{S\text{ geom}}$), the difference in energy between the singlet and triplet states at the optimised triplet geometry ($\Delta E(S-T)_{T\text{ geom}}$) and the difference in energy between the singlet and triplet states at their respective optimised

geometries ($\Delta E(S-T)$). All of these calculated values were found to be less than 0.05 kcal/mol. The RMSD of all Trp and heme atoms between the optimised structures in the singlet and triplet states was 0.00 Å to two decimal places. Based on these calculations, we conclude that the alkylperoxo intermediate has the same geometry and energy on both the singlet and triplet surfaces and that the MECP between these surfaces is located at this geometry.

As the MECP is at the alkylperoxo intermediate and therefore at a minimum on the singlet and triplet surfaces, both states are essentially degenerate in a wide region surrounding their equilibrium geometry. Hence the rate of spin-state change will depend on the magnitudes of the coupling matrix elements between the corresponding wavefunctions. These could be due to spin-orbit coupling or to spin-spin coupling. As the two wavefunctions essentially involve occupation of the two same spatial orbitals by two unpaired electrons, with spins either parallel or antiparallel, the spin-orbit coupling contribution is likely to be small. However, spin-spin coupling at iron is large.²²² We have found previously that coupling matrix elements of as small as 1 cm⁻¹ are sufficient to impart net transition probabilities of ca. 0.1%, which would yield a rate constant for spin crossover in the present case of the order of 10⁹ s⁻¹.^{139,211,223} This is much larger than the rate constants expected for crossing TSII, and indeed TSI, starting from the alkylperoxo intermediate, and hence spin-state change should not be the bottleneck to reaction.

9.3.4. Calculated barriers for all systems

Figure 64 Shows multiple pathways for the reaction of IDO with L-Trp, each starting from a different structure taken from MD simulations. All profiles show the same structure in qualitative terms, although the range in barrier heights is large (around 10 kcal/mol for both barriers). Table 22 summarises the statistical properties of the set of calculated barrier heights for each system. Figure 68 defines $\Delta E(TSI)$, $\Delta E(TSII)$ and other relative energies. As TSII is the highest point on each reaction profile relative to the Michaelis complex, it is $\Delta E(TSII)$ (or rather the exponential average of these values for each system) that we will compare to the activation free energy inferred from the

experimental k_{cat} s. Table 23 summarizes the exponentially averaged calculated barrier heights and experimental kinetic data for all systems.

The ranges and standard deviations for each set of calculated barrier heights are fairly large relative to the differences in barrier heights between systems (for both exponentially averaged calculated barrier heights and for barrier heights derived from experimental k_{cat} s via the Eyring equation). The number of pathways is also fairly modest for each system. For these reasons, it appears that our current best estimates of $\Delta E(\text{TSII})$ are not sufficiently well converged for us to be able to confidently rank the systems in terms of their reaction rate constants. It is therefore not surprising that no correlation is observed when the exponentially averaged $\Delta E(\text{TSII})$ s are plotted against the log of the experimental k_{cat} s (Figure 69).

If the systems are divided into three groups, a trend does emerge. Group 1 contains IDO with L and D-Trp and TDO with L-Trp. All of these systems have reasonably large k_{cat} s, which are so similar that, given the error bars of the methods we are using, we cannot reasonable expect our calculations to correctly rank them. The exponentially averaged $\Delta E(\text{TSII})$ s for these systems are between 16.7 kcal/mol and 18.2 kcal/mol. Group 2 contains IDO with 1-Me-L-Trp, which has a significantly smaller k_{cat} and a slightly higher exponentially averaged $\Delta E(\text{TSII})$ of 19.1 kcal/mol. Group 3 contains TDO with D-Trp and 1-Me-L-Trp. These systems were found experimentally not to turnover at a detectable rate. This is in agreement with the higher exponentially averaged $\Delta E(\text{TSII})$ s for these systems of 20.6 kcal/mol and 21.8 kcal/mol.

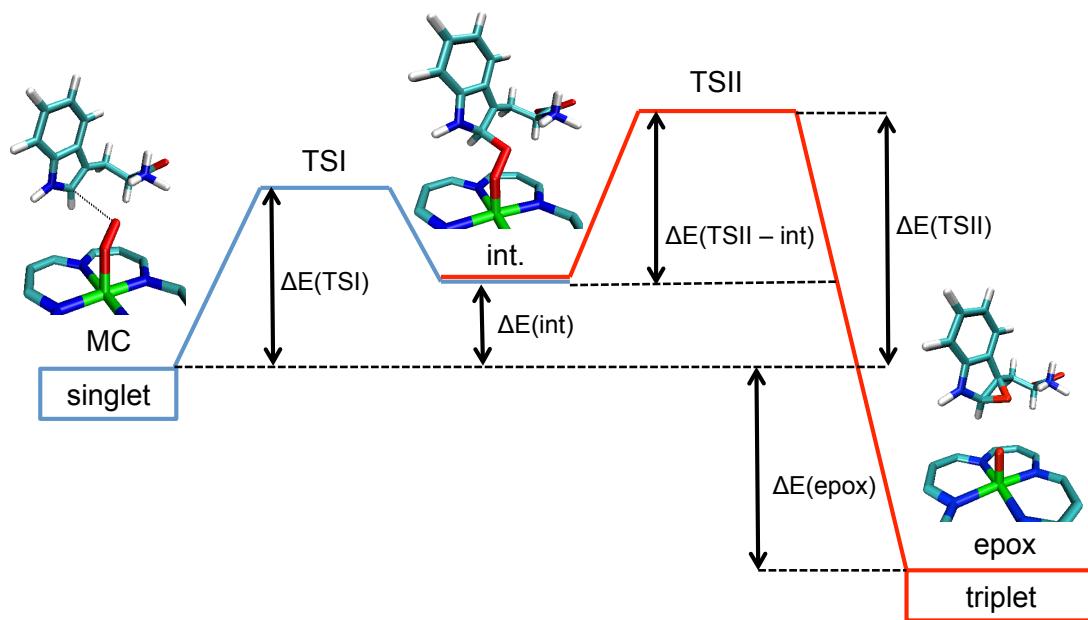


Figure 68 Scheme of the first two steps in the hemedioxygenase reaction. The key differences in energy between the various stationary points and the Michaelis complex are shown.

Table 22 Minimum, maximum, mean and standard deviations of the first (a) and second (b) barrier heights from QM/MM adiabatic mapping. All values are in units of kcal/mol (except the number of pathways).

a)	$\Delta E(TSI)$				number of pathways
	min	max	mean	SD	
IDO Trp	11.7	22.6	16.0	3.2	11
IDO 1-Me-L-Trp	14.7	20.9	18.2	2.1	9
IDO D-Trp	14.0	19.8	16.8	1.6	11
TDO L-Trp	12.5	17.4	14.7	1.5	10
TDO 1-Me-L-Trp	16.3	26.4	20.8	3.3	8
TDO D-Trp	15.8	24.0	18.9	2.8	8

b)	$\Delta E(TSII)$				number of pathways
	min	max	mean	SD	
IDO Trp	15.5	26.0	20.0	3.2	11
IDO 1-Me-L-Trp	17.9	24.7	21.1	2.2	9
IDO D-Trp	16.9	21.7	19.9	1.6	11
TDO L-Trp	17.4	21.5	19.2	1.4	10
TDO 1-Me-L-Trp	19.4	25.6	23.1	2.2	8
TDO D-Trp	20.7	26.1	23.4	1.7	8

Table 23 Exponentially averaged barrier heights from adiabatic mapping and kinetic data from experiment. “-“ indicates that the values are not available from experiment because the system did not turn over at a detectable rate. The ref. column the source of the experimental data: a¹⁶⁹, b¹⁶⁸, c¹⁷⁰ and d²²⁴.

	ΔE(TSI) kcal/mol	ΔE(TSII) kcal/mol	exp. kcat (s ⁻¹)	exp. ΔG [‡] (Kcal/mol)	exp. KM (μM)	ref.
IDO Trp	13.0	16.7	1.4	17.2	7	a
IDO 1-Me-L-Trp	15.9	19.1	0.027	19.6	150	b
IDO D-Trp	15.3	18.2	3.93	16.6	1570	a
TDO L-Trp	13.6	18.2	19.5	15.7	144	c
TDO 1-Me-L-Trp	17.5	20.6	-	-	-	b
TDO D-Trp	16.8	21.8	0	-	16500	d
TDO-H55A						
L-Trp	-	-	2.86	16.8	133	c
TDO-H55A						
1-Me-L-Trp	-	-	0.048	19.2	59	b

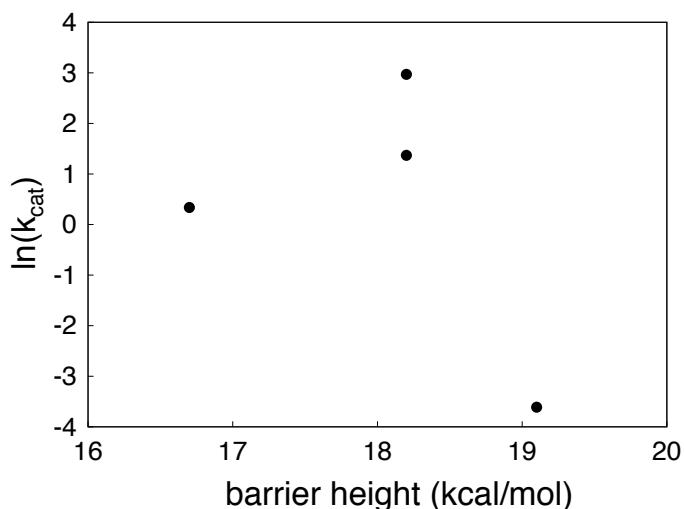


Figure 69 Logarithm of experimental rate constant plotted against exponentially averaged calculated ΔE(TSII) barrier heights (for the four systems where values for both properties are available).

9.3.5. Correlating Michaelis complex geometry with barrier heights

As discussed in section 9.1, previous work in our group and other groups has shown that barrier heights obtained by adiabatic mapping are often highly sensitive to

even seemingly minor variations in the starting structure used. Here we have attempted to find quantitative relationships between the starting structures used in this study and the barrier heights obtained.

Before looking into structural correlations it is useful to look at how the different relative energies shown in Figure 68 correlate with each other. Table 24 is a correlation matrix showing the gradient of the line of best fit (from linear regression) and the correlation coefficient (R^2) for each pair of ΔE s. $\Delta E(\text{TSI})$, $\Delta E(\text{int})$ and $\Delta E(\text{TSII})$ are all highly correlated, implying that the physical factors which affect the magnitude of one of these ΔE s will most likely affect the magnitude of the others. Surprisingly, $\Delta E(\text{TSII} - \text{int})$ does not correlate at all with the overall barrier to the reaction, $\Delta E(\text{TSII})$. This may be related to the observation that $\Delta E(\text{TSII} - \text{int})$ is slightly negatively correlated with $\Delta E(\text{int})$. The effect of reductions in $\Delta E(\text{TSII} - \text{int})$ on the magnitude of $\Delta E(\text{TSII})$ may be offset by increases in $\Delta E(\text{int})$. $\Delta E(\text{TSII})$ is somewhat correlated with $\Delta E(\text{epox})$, so again, the physical factors which affect the magnitude of one may also affect the other.

Table 24 Correlation matrix of the ΔE s illustrated in Figure 68. Cells in blue show the gradient of the line of best fit from linear regression and the cells in white show the R^2 correlation coefficients. The entire data set of energies from all pathways and all systems was used.

	$\Delta E(\text{TSI})$	$\Delta E(\text{int})$	$\Delta E(\text{TSII} - \text{int})$	$\Delta E(\text{TSII})$	$\Delta E(\text{epox})$
$\Delta E(\text{TSI})$	1.00	0.82	-0.05	0.77	-0.04
	1.00	0.73	0.02	0.84	0.00
$\Delta E(\text{int})$	0.89	1.00	-0.20	0.80	-0.19
	0.73	1.00	0.24	0.84	0.05
$\Delta E(\text{TSII} - \text{int})$	-0.35	-1.23	1.00	-0.23	0.88
	0.02	0.24	1.00	0.01	0.17
$\Delta E(\text{TSII})$	1.09	1.05	-0.05	1.00	-0.06
	0.84	0.84	0.01	1.00	0.00
$\Delta E(\text{epox})$	-0.07	-0.26	0.19	-0.06	1.00
	0.00	0.05	0.17	0.00	1.00

Various structural parameters of the Michaelis complexes were examined for each pathway (from every system) and correlations with barrier heights were sought. Analysis of the Michaelis complex geometry was performed on structures reoptimised without restraints starting from the reverse adiabatic mapping structure with a $R(\text{Ot-C2})$ value of 2.8 Å.

A number of single geometric parameters that were found to correlate with barrier heights $\Delta E(\text{TSI})$ and $\Delta E(\text{TSII})$ are reported in Table 25. Many other geometrical

parameters were investigated but showed no strong correlation with the barrier heights. A select few of these are reported in

Table 26. Combinations of single parameters were investigated in an attempt to find stronger correlations. These are reported in Table 27. All of the atoms involved in the various geometrical parameters are labelled in Figure 70.

The distance R(Ot-C2) was the single parameter that correlated most strongly with both barrier heights $\Delta E(TSI)$ and $\Delta E(TSII)$. The only parameter involving a protein residue that was found to correlate with barrier height was R(Ot-HN). The deviation of the angle A(Op-Ot-C2) from 99° was found to correlate modestly with barrier height. The angle 99° was chosen simply because this resulted in the strongest correlation.

The angles of intersection between sets of two planes were also examined. Each of these planes was defined by three points. C2, C3 and C4 defined the Trp plane, Fe, Op and Ot defined the O2 plane and NA, NB and NC defined the heme plane. For the Trp-heme pair, the deviation of the angle of intersection from 90° was plotted against the barrier heights. For the Trp-O2 pair, the deviation of the angle of intersection from 0° was plotted. These plane intersections were both found to correlate weakly with $\Delta E(TSI)$ and slightly more strongly with $\Delta E(TSII)$.

The distance R(Ot-N) was not found to correlate at all with barrier height. The combination of distances (Ot-C2) + (Ot-N) correlated less well with barrier height than R(Ot-C2) on its own. This shows that using these two distances together as selection criteria for choosing adiabatic mapping starting structures was a fairly poor choice.

The combination of single geometric parameters that gave the strongest correlation with barrier height $\Delta E(TSI)$ was found to be (Ot-C2) + 0.33(Op-C2). This also correlated strongly with $\Delta E(TSII)$. That which gave the strongest correlation with barrier height $\Delta E(TSII)$ was (Ot-C2) + 0.33(Op-C2) + 0.05(HN - Ot). However, as the contribution on R(HN - Ot) is very small and the improvement in correlation by adding this contribution is modest, we have focused only on the parameter (Ot-C2) + 0.33(Op-C2) in our further analysis.

Figure 71 shows plots of the value of (Ot-C2) + 0.33(Op-C2) for each Michaelis complex structure against the resultant barrier heights. The distributions of both energies and values of (Ot-C2) + 0.33(Op-C2) for the different systems are wide and overlap with each other to a high degree. Because of this, it is difficult to find any meaningful

interpretation of the distributions for each system relative to the others. It is notable however, that the minimum values of $(\text{Ot-C}2) + 0.33(\text{Op-C}2)$ and of both barrier heights are higher for TDO with 1-Me-L-Trp and TDO with D-Trp than for any of the other systems.

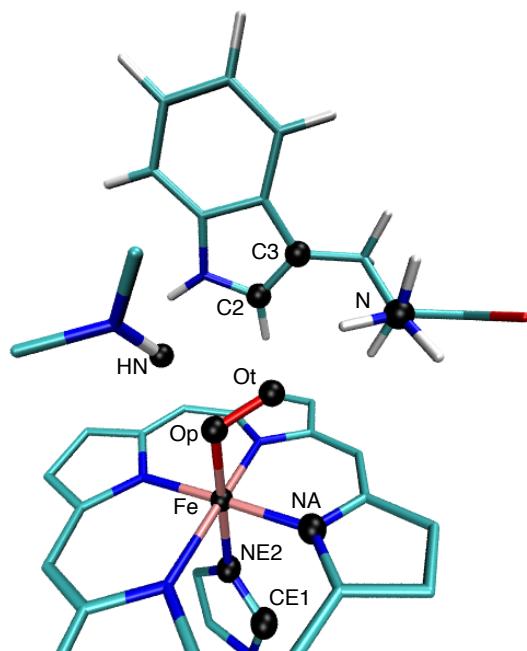


Figure 70 Atoms involved in the structural parameters investigated for correlation with barrier heights.

Table 25 Correlation of single geometric parameters of the Michaelis complex with barrier heights $\Delta E(TSI)$ and $\Delta E(TSII)$. The gradient of the line of best fit from linear regression is shown in blue, R^2 correlation coefficients are shown in white. Only the parameters which were found to correlate with barrier heights are shown here.

	$\Delta E(TSI)$	$\Delta E(TSII)$
R(Op-C2)	12.28	10.15
	0.37	0.36
R(Ot-C2)	42.66	30.75
	0.50	0.37
R(Ot-HN)	1.85	2.23
	0.15	0.30
R(Op-C3)	15.14	10.14
	0.44	0.28
A(Op-Ot-C2 dev.)	0.17	0.15
	0.20	0.23
Trp-O2 plane	0.09	0.09
	0.11	0.19
R(Ot-C3)	11.84	6.65
	0.20	0.09
Trp-Heme plane	0.14	0.19
	0.05	0.14

Table 26 Correlation of single geometric parameters of the Michaelis complex with barrier heights $\Delta E(TSI)$ and $\Delta E(TSII)$. The gradient of the line of best fit from linear regression is shown in blue, R^2 correlation coefficients are shown in white. Only the parameters which were found not to correlate with barrier heights are shown here.

	$\Delta E(TSI)$	$\Delta E(TSII)$
D(Op-Ot-C2-C3)	0.05	0.00
	0.05	0.00
A(C3-C2-Ot)	0.14	0.04
	0.04	0.01
R(Ot-N)	3.50	-2.39
	0.00	0.00
D(CE1-NE2-Fe-NA)	0.00	0.00
	0.00	0.00

Table 27 Correlation of combinations of single geometric parameters of the Michaelis complex with barrier heights $\Delta E(TSI)$ and $\Delta E(TSII)$. The gradient of the line of best fit from linear regression is shown in blue, R^2 correlation coefficients are shown in white.

	$\Delta E(TSI)$	$\Delta E(TSII)$	parameters
(Ot-C2) + (Ot-N)	20.50	12.17	
	0.26	0.13	
(Op-C2) + (Ot-C2)	13.34	10.63	
	0.56	0.50	
(Ot-C2) + $\alpha(Op-C2)$	32.05	24.66	$\alpha = 0.33$
	0.70	0.58	
(Ot-C2) + $\alpha(Op-C2)$ + $\beta(HN - Ot)$	26.17	21.91	$\alpha = 0.33$
	0.67	0.67	$\beta = 0.05$

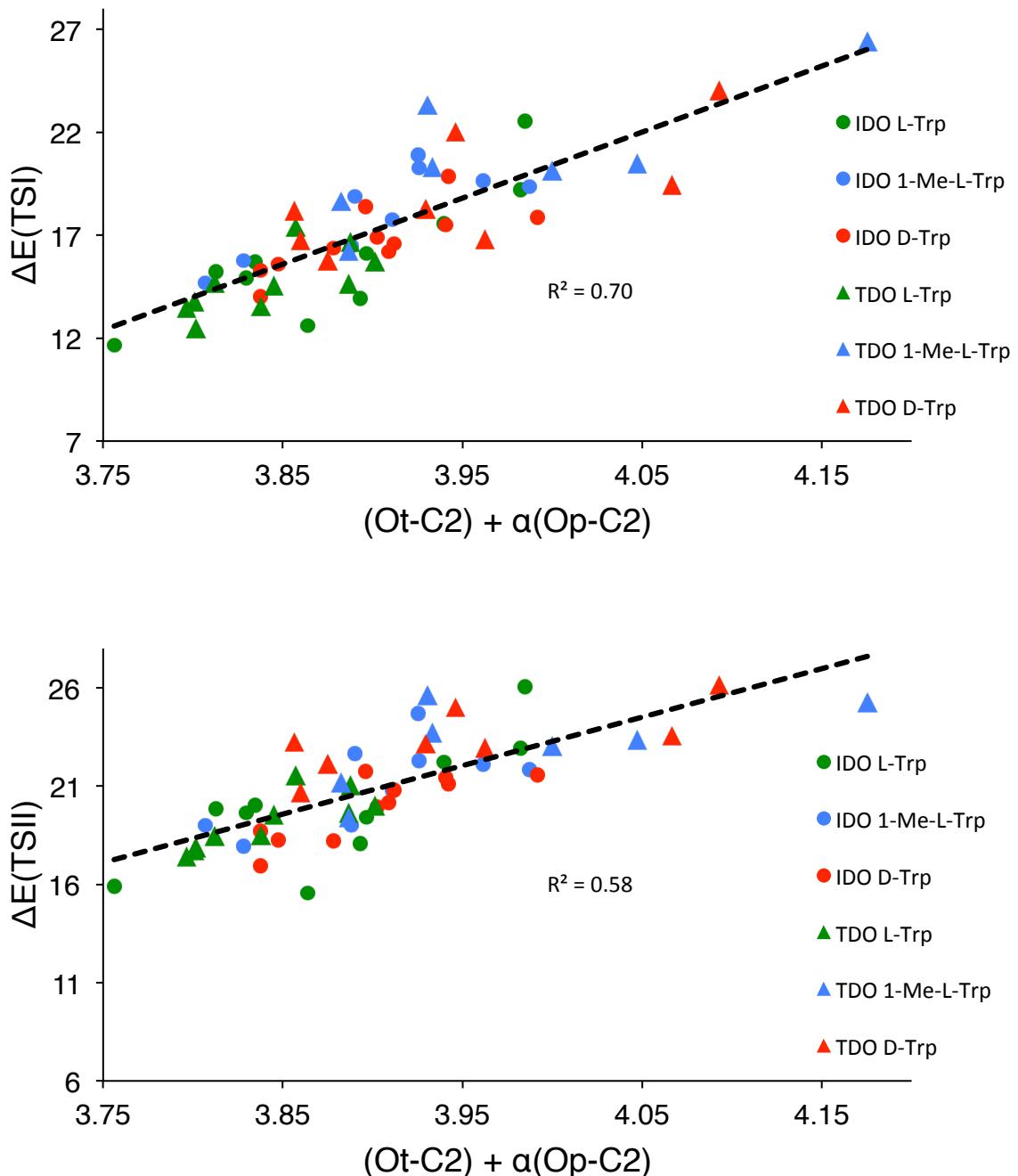


Figure 71 The values of $(Ot\text{-}C2) + 0.33(Op\text{-}C2)$ for each Michaelis complex plotted against the barrier heights $\Delta E(TSI)$ (top) and $\Delta E(TSII)$ (bottom).

9.3.6. Effect of enzyme environment on barrier height

To investigate the effect that the enzyme environment has on barrier height, single point energy calculations were performed on the QM region alone in the gas phase. The structures used for this were extracted from the stationary points along the QM/MM adiabatic mapping pathways. This was done only for the 11 pathways calculated for IDO with L-Trp. Table 28 reports the differences in the ΔE s defined in Figure 68 between the QM/MM and gas phase energies.

The inclusion of the enzyme environment has almost no significant effect on either the mean or exponentially averaged barrier heights for the first two steps. This suggests that the major role of the enzyme is to bind the substrate in the correct orientation relative to the oxyheme group rather than to provide any electrostatic stabilisation of the transition state. The enzyme may also help to stabilize the oxyheme reactive species, protect the system from side-reactions, and so on.

As there was some correlation between the distance $R(Ot-HN)$ and barrier height, similar calculations were performed to study the role of Ala264. Single point QM/MM calculations were performed on structures extracted from the stationary points along the QM/MM adiabatic mapping pathways on IDO with L-Trp. The same QM region was the same as that used throughout this work. The only atoms included in the MM region were those belonging to Ala264.

Table 29 summarises the difference that adding Ala264 has on the gas phase ΔE s. The effect on the barrier heights is again negligible. Ala264 does however seem to stabilise the epoxide + compound II intermediate.

*Table 28 Effect of full enzyme environment on relative energies. Enzyme energies (enz.) are from QM/MM, gas phase energies are from QM single point calculations on the same structures but with the entire MM region removed. All energies calculated with LACV3P**++ basis set. Energies are in kcal/mol.*

	TSI	int	TSII - int	TSII	epox
mean $\Delta\Delta E$ (enz. - gas)	1.2	0.9	-0.2	0.7	-1.7
enz. exp. av.	13.0			16.7	
gas exp. av.	12.7			16.9	
Δ exp. av. (enz. - gas)	0.3			-0.3	

*Table 29 Effect of Ala264 on relative energies. “ala.” denotes QM/MM calculations with the same QM region as used throughout this study and only Ala264 in the MM region. All energies calculated with LACV3P**++ basis set. Energies are in kcal/mol.*

	TSI	int	TSII - int	TSII	epox
mean $\Delta\Delta E$ (ala. - gas)	0.9	-1.4	0.7	-0.7	-2.3
ala. exp. av.	13.3			16.9	
ala. exp. av.	12.7			16.9	
Δ exp. av. (ala. - gas.)	0.6			-0.1	

9.3.7. Effect of empirical dispersion correction

The effect of adding Grimme’s empirical dispersion correction to DFT in gas phase QM and enzyme QM/MM calculations has been investigated previously by our group^{225,226} as well as many others. In the studies from our group, it was found that the inclusion of dispersion correction significantly improved the accuracy of the calculated QM and QM/MM barrier heights. The effect on the geometry was large for gas phase QM systems but small for QM/MM systems.

Table 30 shows the effect of removing the dispersion component from the total QM/MM energies. We also looked at the effect of dispersion correction on geometry by reoptimising the minima along one pathway without dispersion (Table 30). Here we see similar results to those previously reported. The average effect of dispersion on the mean and exponentially averaged overall barrier height is significant (a reduction of more than 2 kcal/mol), while the effect on the geometry is slight (Figure 72). The dispersion-free energies are given for structures optimised with dispersion, however as the effect on geometry is slight, this is unlikely to cause significant error in our estimate of the effect of dispersion.

The direction of the effect of dispersion on the relative energies of the stationary points is as expected. The substrate is drawn closer to the heme group in both TSI and the intermediate relative to the Michaelis complex, so the dispersion interactions are expected to be stronger thus stabilising these structures. The substrate group is pushed away from the heme group in TSII relative to the intermediate, so adding dispersion is expected to destabilise TSII relative to the intermediate and increase ΔE (TSII - int), and this is indeed observed.

Our best estimates of the barrier to reaction are in good agreement with experimental rate constants (see Table 23). If dispersion were not included, our calculated barriers would be higher than predicted from experiment.

Table 30 Effect of empirical dispersion correction on relative energies. Energies are in in kcal/mol.

	TSI	int	TSII - int	TSII	epox
mean $\Delta\Delta E$ (disp. - nodisp.)	-2.9	-2.7	0.3	-2.4	0.7
disp. exp. av.	13.0			16.7	
nodisp. exp. av.	15.2			18.9	
Δ exp. av. (disp. - nodisp.)	-2.2			-2.2	

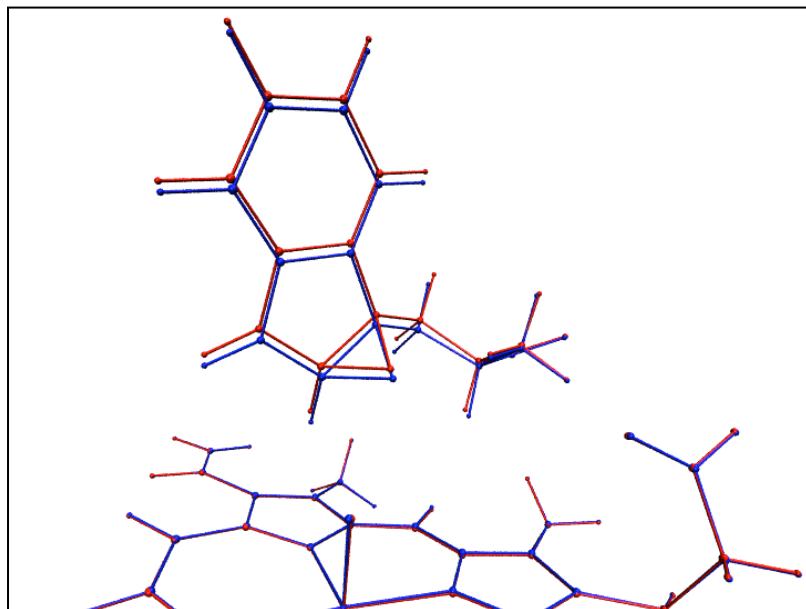


Figure 72 Effect of dispersion on geometry. The epoxide and CpII structure optimised by QM/MM with dispersion (blue) and without dispersion (red). The differences in geometry were even more slight for the other minima along the reaction pathway.

9.3.8. Ruled out mechanisms

For many years, the leading mechanisms in the literature for the heme dioxygenase reactions involved deprotonation of the indole N-H hydrogen of tryptophan by an active site base as the first step. This was until recent experimental and computational studies showed that this proton transfer is unlikely to occur.^{168,170,177,208}

To further test these mechanism, taking a QM/MM optimised Michaelis complex of TDO with Trp, we removed the indole N-H proton from tryptophan and added it to the His55 NE2 nitrogen (with which it was hydrogen bonded). The proton was added in plane with the histidine imidazole and with an N-H bond length of 1.01 Å (the same as the other imidazole N-H bond length). The resulting indole N-H distance was 2.59 Å. A closed shell singlet QM/MM optimisation with no restraints was then performed with only the imidazole and tryptophan in the QM region. The proton spontaneously transferred back to the indole nitrogen, indicating that the proton-transferred state is not stable, adding further evidence against these mechanisms.

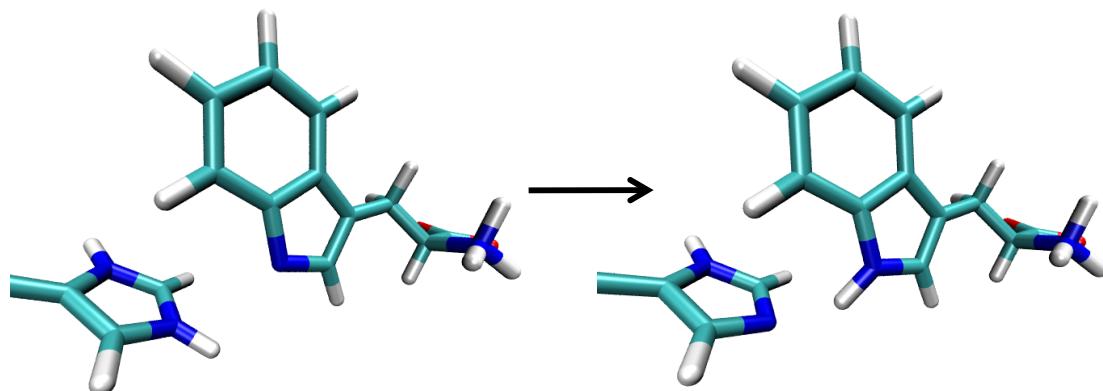


Figure 73 In this figure (left), the indole N-H proton has been removed and placed on the His55 NE2 nitrogen. During a QM/MM optimisation of this structure, the proton spontaneously transferred back to the indole N-H nitrogen (right).

Some other mechanisms require the Michaelis complex to resemble the Pauling model of a closed shell singlet state for oxy-heme.^{168,206} We found the closed shell singlet state to lie 19 kcal/mol above the open shell singlet ground state. As this is higher than our best estimate for the barrier height for the reaction of IDO with Trp, any mechanism involving a closed shell singlet Michaelis is inconsistent with our work. As already noted in 9.3.1 however, it is known that DFT does have significant deficiencies in its ability to accurately model oxyheme systems.

Although we did not perform any full adiabatic mapping of the reaction in the closed shell singlet state, we did perform single point calculations on the stationary points from the open shell singlet and triplet surfaces as reported in Figure 74. These numbers are likely to have error bars of at least a few kcal/mol associated with them, as they are

not derived from calculations in which the structure has been optimised on the closed shell surface. The results suggest that the alkylperoxy intermediate structure is not a minimum on the closed shell singlet surface and may even be a transition state.

In a previous QM/MM study of IDO and TDO, the alkylperoxy species was thought to be a transition state directly linking the Michaelis complex and the epoxide + CpII state for both enzymes.²⁰⁸ This study found the ground state of TDO to be a closed shell singlet, yet the ground state of IDO to be an open shell singlet. On the basis of the extremely high level of similarity between the electronic structures of IDO and TDO shown in this work, we believe this inconsistency to be a result of the group's failure to converge to the correct ground state for the Michaelis complex of TDO. In light of our crude estimate of the reaction profile for the closed shell singlet shown in Figure 74, it seems possible that the group's (proposed) failure to converge to the correct broken-symmetry open shell states may also explain why they found the alkylperoxy species to be a transition state where we have found it to be an intermediate.

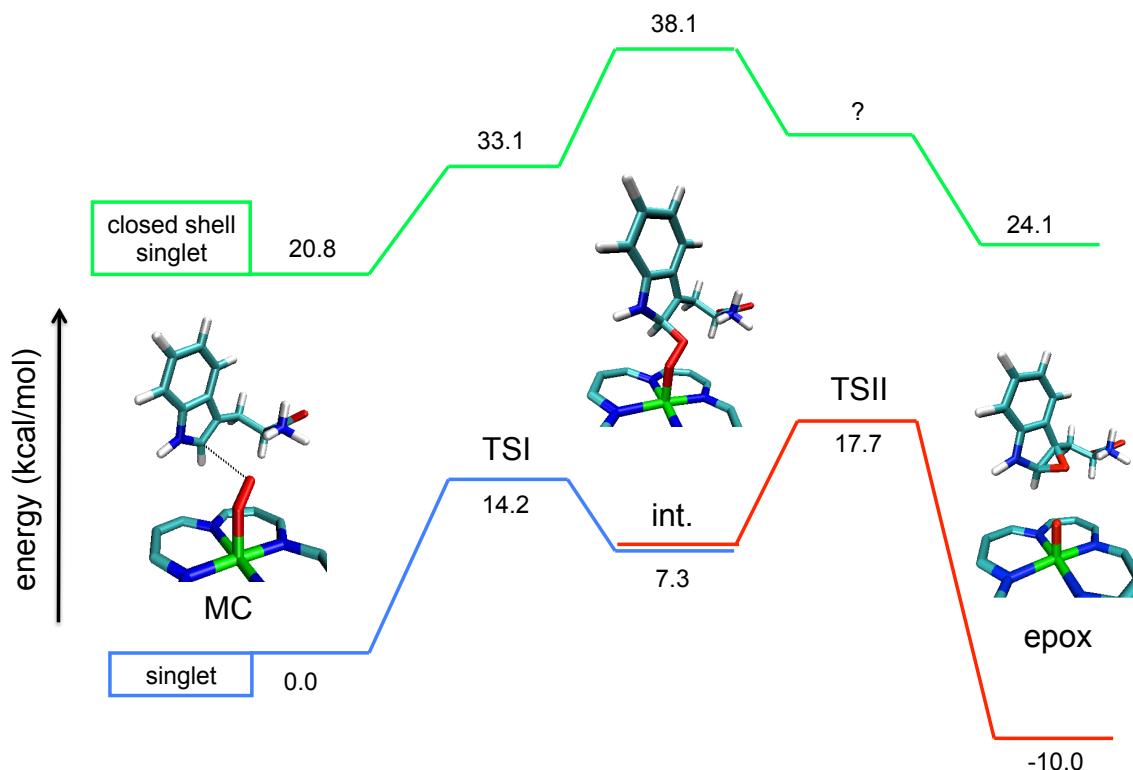


Figure 74 A crude estimate of the closed shell singlet reaction pathway was calculated by performing single point energy calculations on the stationary points from the singlet and triplet surfaces of one IDO with Trp pathway. The calculation for TSII did not converge.

Radical addition to the C3 carbon was also tested by performing adiabatic mapping with R(Ot-C3) as the reaction coordinate. This was done for a single pathway only using RODFT in the triplet state. It should be noted also that the structure used for this calculation was selected due to a short Ot-C2 and Ot-NH₃⁺ nitrogen distance, not a short Ot-C3 distance. Despite these methodological deficiencies we are confident that this mechanism can be ruled out because the resulting energy profile rose so steeply in energy and clearly produced no minima (Figure 75).

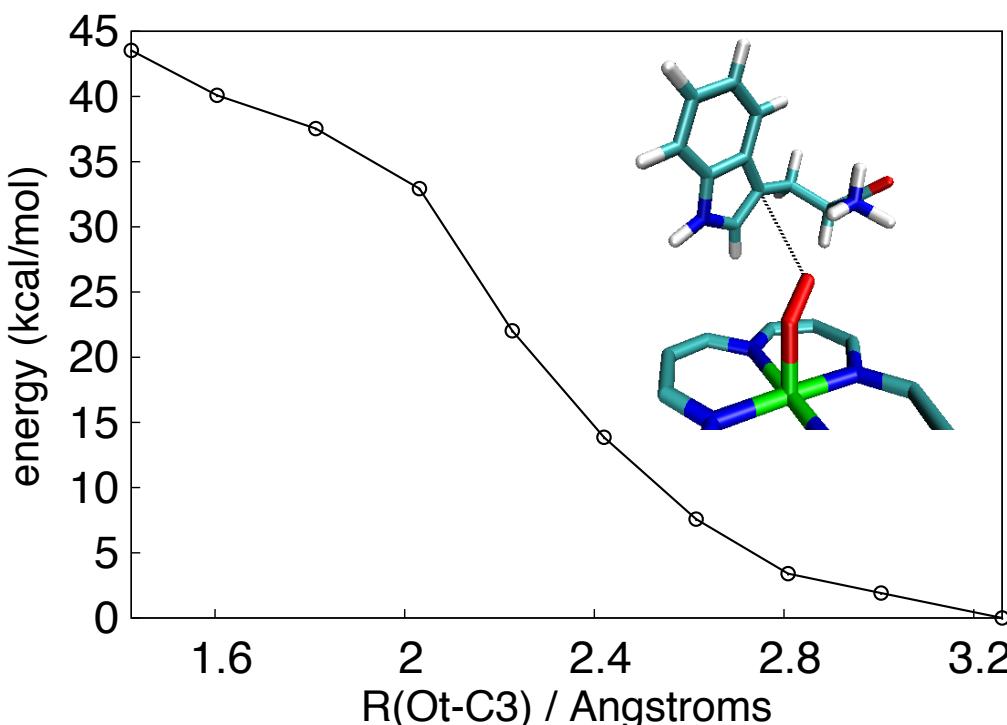


Figure 75 Adiabatic mapping was performed using the distance $R(Ot-C3)$ as the reaction coordinate to investigate oxygen addition to the C3 carbon as the first step. The energy rose steeply along this pathway and produced no minimum.

Chapter 10. Heme dioxygenase conclusions

10.1. Conclusions

Indoleamine 2,3-dioxygenase (IDO) and tryptophan 2,3-dioxygenase (TDO) belong to a class of enzymes called the heme dioxygenases, which catalyse the first step in the metabolism of L-tryptophan (L-Trp), incorporating both atoms of molecular oxygen into the substrate to form N-formyl-L-kynurenine. Due to their implication in a number of diseases, including a variety of cancers, both enzymes have received considerable interest as therapeutic drug targets.^{180,227} Much work has been undertaken in search of novel, effective heme dioxygenase inhibitors, some of which have already been the subject of clinical trials.¹⁵⁹ Modern regulations and general best practice in the pharmaceutical field requires that drugs are highly specific to their target in order to avoid potentially harmful side effects. Given the high level of similarity between IDO and TDO in terms of reactivity and active site structure, designing drugs that are specific to one of these enzymes only could pose a significant challenge. Consequently, detailed insight into the similarities and subtle differences in substrate binding, inhibition and reactivity between these enzymes is likely to prove highly valuable.

With this motivation, molecular modelling has been used in the present work to study the reactivity and binding of human IDO (hIDO), a bacterial TDO (XcTDO) and the H55A mutant XcTDO with L-Trp, D-Trp and 1-Me-L-Trp. Molecular dynamics simulations, binding free energy calculations and reaction modelling with QM/MM adiabatic mapping have been performed in an attempt to reproduce and rationalise experimentally observed differences in binding and reactivity of these systems.

The crystal structure for XcTDO with L-Trp bound in the active site revealed a number of substrate-protein hydrogen bonding and salt bridge interactions. Comparison with the substrate free crystal structure showed that substrate binding is associated with several protein conformational changes. One of these changes is the movement of Arg117 away from the solvent in towards the active site to form a salt bridge with the carboxylate group of the substrate. A loop section near the active site is missing from the substrate

free crystal structure due to disordered electron density. Upon substrate binding this section becomes ordered and one of its residues (Thr254) hydrogen bonds with the substrate NH₃⁺ group. As O₂ was missing from the crystal structure of XcTDO, its interactions with the substrate and protein were not known for certain. The MD simulations of L-Trp showed that the terminal oxygen of O₂ forms hydrogen bonds with the substrate NH₃⁺ group and a carbonyl oxygen from the protein backbone. This is in agreement with evidence from an ENDOR spectroscopy study¹⁶⁵ but inconsistent with previous suggestions that bound O₂ hydrogen bonds with and abstracts the substrate indole N-H proton.²⁰⁶

None of the available crystal structures for hIDO have substrates or close substrate analogues bound, so little was known for certain about substrate binding in IDO. The MD simulations revealed very similar substrate binding interactions between the two enzymes with the major difference being the lack of an IDO equivalent to the substrate indole N-H to His55 hydrogen bond found in TDO. As was observed in TDO by comparing the substrate bound and unbound structures, upon substrate binding, an active site arginine (Arg231) in IDO moved from a solvent facing position in towards the active site to form a salt bridge with the substrate NH₃⁺ group. From this observation it is concluded that substrate binding proceeds by an induced fit mechanism (at least to some extent). The MD simulations indicate that the substrate binding interactions are largely the same for all substrates and all enzymes studied (except of course where some interactions were impossible, for example due to the lack of an indole N-H group in 1-Me-L-Trp). The most notable exception to this was TDO with D-Trp in which there was significantly reduced hydrogen bonding and salt bridge interaction between the substrate and protein. This is consistent with the high inhibitor constant (K_i) and lack of activity for this system. In addition, TDO with 1-Me-L-Trp (which also does not turn over) was found to have somewhat reduced salt bridge interactions.

XcTDO has four active sites, whereas hIDO has only one. Binding free energies were calculated by running MM/PBSA calculations on each active site separately using structures taken from individual production runs, producing a set of free energies for each system. The calculated binding free energies had internally converged for each individual active site of a given production run, however they had clearly not converged between different active sites and production runs. Plotting the calculated binding free energies against the average number of substrate binding interactions for the corresponding structures revealed a strong negative correlation (the greater the average number of

interactions, the lower the binding free energy). This showed that the calculated binding free energies were dominated by the electrostatic substrate binding interactions detailed above.

The wide spread of average numbers of substrate binding interactions between active sites and production runs for the same systems suggested that the lack of convergence in the calculated binding free energies was related to poor convergence with respect to active site structure. In particular, it was observed that whether a particular trajectory produced a high or low calculated binding free energy depended on whether a salt bridge between the active site arginine and the carboxylate group of the substrate had been formed. Given the poor overall convergence, all further analysis of the MD simulations was performed on the trajectories that produced the lowest calculated binding free energies on the assumption that they best represented the equilibrated ensembles of structures. Despite the lack of convergence, there was a trend in the binding free energies of the TDO systems. Those which are known from experiment to turn over had low or similar relative binding free energies, whereas those which do not turn over had higher binding free energies. In particular D-Trp had a much higher binding free energy, consistent with its high experimental K_i and evidence from MD of its reduced substrate-protein binding interactions.

Reaction modelling with QM/MM was entirely consistent with the calculations and proposed mechanism of the Morokuma group.¹⁷⁷ The results of adiabatic mapping indicated that the reaction proceeds by the same mechanism in both enzymes. The ground state of the Michaelis complex was found to be an open shell singlet, with one unpaired electron centred on the iron and the other on O₂, supporting the Weiss ferric superoxide model of oxy-heme. A second singlet state and the two equivalent triplet states (differing from the singlet states only by the spin orientation of a single electron) were found to lie within 3.5 kcal/mol of the ground state. Radical addition to the C2 carbon was found to produce an alkylperoxo intermediate, in agreement with Morokuma but in contrast to another QM/MM study which found this species to be a transition state.²⁰⁸ Homolytic cleavage of the O-O bond led to the formation of the ferryl-oxo heme species known as compound II and an epoxide intermediate. This is in agreement with the detection of compound II during turnover of IDO by Raman spectroscopy.¹⁷⁴ The first step was found to have a lower barrier on the singlet surface than on the triplet surface. An investigation into the overlap between the singlet and triplet surfaces in the region of the alkylperoxo intermediate led to the conclusion that the minimum energy crossing point (MECP) is

located at this structure. Given this and the fact that the second step has a much lower barrier on the triplet surface, it is concluded that spin state change is most likely to occur at the alkylperoxo intermediate. The rate of spin state change is estimated to be fast relative to the rate of homolytic cleavage of the O-O bond, so the rate of spin state change is not thought to be a bottleneck to reaction.

Approximately ten reaction profiles were calculated for each system starting from a range of structures taken from MD simulations. Exponential averaging was used to produce a best estimate of the barrier to reaction for each system. This is a more thorough procedure than was used by any of the previous QM/MM studies on the heme dioxygenases, which involved the calculation of few¹⁷⁷ or even single pathways.^{176,208} For all systems (that turned over experimentally), the calculated barrier heights were within 2.5 kcal/mol of the experimental free energy barriers (calculated from k_{cat} s via the Eyring equation). Relatively high standard deviations for the barrier heights from individual pathways for a given system indicated that the exponentially averaged barrier heights were not tightly converged. Due to this and possibly other methodological errors, the exponentially averaged barriers were not able to correctly rank every system in terms of their experimental k_{cat} s. However, by grouping the systems in terms of activity, a trend did emerge. The group of systems with the highest k_{cat} s (IDO with L-Trp, IDO with D-Trp and TDO with L-Trp) had the lowest barriers (16.7-18.2 kcal/mol). IDO with 1-Me-L-Trp, which has a low k_{cat} , had a slightly higher calculated barrier (19.5 kcal/mol). The other two systems studied, which are known experimentally not to turn over (TDO with D-Trp and 1-Me-L-Trp), had the highest calculated barriers (20.6 and 21.8 kcal/mol).

Although evidence has already mounted against the earlier proposed reaction mechanisms^{168,206}, several of these were investigated here anyway. Radical addition to the C3 carbon was ruled out on the basis that a pathway calculated for this process rose steeply in energy without producing a minimum. To test the mechanisms that require proton abstraction as an initial step, the substrate indole N-H proton was removed and placed on the NE2 nitrogen of His55 (with which it was hydrogen bonded). Upon optimisation of this structure, the proton spontaneously transferred back to the indole nitrogen, providing further evidence against these mechanisms. The closed shell singlet state of the Michaelis complex was found to lie 19 kcal/mol above the ground state, suggesting that mechanisms involving this state is energetically unfeasible.

The use of Grimme's empirical dispersion correction was found to reduce the calculated barrier heights by around 2 kcal/mol, improving the agreement with experiment. Dispersion was found to have a negligible effect on the geometries. This is consistent with previous experience in the group employing empirical dispersion correction to QM/MM calculations.²²⁸

Many enzymes are thought to catalyse reactions by electrostatic stabilisation of the transition state.²²⁹ It has even been claimed by recent Nobel laureate Arieh Warshel that: "such contributions appear to account for the catalytic effects of all of the enzymes that were examined by consistent computational studies".²³⁰ To test for this effect in the heme dioxygenases, the energies of the stationary points along the QM/MM adiabatic mapping pathways were recalculated with single point calculations on the QM region in the absence of the MM environment. The removal of the MM region changed both the mean and exponentially averaged barrier heights by less than 1 kcal/mol, leading to the conclusion that electrostatic stabilisation of the transition state is not a catalytic effect in the heme dioxygenase enzymes.

In an attempt to understand why there is such a large range of barrier heights from individual adiabatic mapping pathways for a given system, correlations were sought between the minimum geometry of the Michaelis complex and barrier height. For this analysis, all pathways from all systems were used (a total of 57). A strong correlation was found between barrier height and the position of the C2 carbon with respect to both the terminal and proximal oxygens of heme bound O₂ at the Michaelis complex. The R² correlation coefficient was maximized by correlating the energy with the combined parameter (Ot-C2) + 0.33(Op-C2), which is referred to as λ . The associated line of best fit for this correlation had the formula $\Delta E = 24.7\lambda + 75$. As well as giving insight into the variation in barrier heights obtained with different starting structures, this led to the hypothesis that the differences in observed activity of these enzymes could be related to variation in their ability to position the substrates relative to the heme bound O₂ for optimal reactivity.

To test this theory, the MD simulations of all systems were analysed and histograms of λ were constructed. From the formula $\Delta E = 24.7\lambda + 75$, for a pathway to have a barrier of less than 20 kcal/mol (corresponding to a rate of around 0.01 s⁻¹), the Michaelis complex would have to have a value for λ of less than 3.87 Å. All systems that turned over at a reasonable rate were found to spend an appreciable amount of time (more than 2% of the MD simulations) at structures with values of λ lower than 3.87 Å. IDO

with 1-Me-L-Trp, which turns over slowly, rarely fit this criterion (0.24%). The two systems which did not turn over (TDO with D-Trp and 1-Me-L-Trp) were found to almost never have values of λ lower than 3.87 Å (less than 0.05% of structures).

It seems that, due to the high intrinsic oxidising power of the oxy-heme species, electrostatic stabilisation of the transition state is not required in these enzymes. Rather, a major role of enzyme is to bind the substrate in an optimal position relative to heme-bound O₂, ready for radical addition by ferric superoxide. In addition to this, the enzyme will undoubtedly perform other functions, such as protecting the reactive intermediate species from side-reactions, modulating the properties of the various heme species and moderating substrate binding and release.

10.2. Further work

The MD simulations for the systems studies here did not converge to equilibrium ensembles of structures with respect to the substrate-protein binding interactions. This was thought to be the cause of the poor convergence of the calculated binding free energies. It seems likely that longer MD simulations would improve structural convergence. If this is the case, it would be interesting to see if this improves the accuracy of the binding free energies calculated with MM/PBSA.

It appears that the binding free energies calculated with MM/PBSA for the IDO systems are not directly comparable to those calculated for the TDO systems. MM/PBSA's ability to predict absolute (as opposed to relative) binding free energies is known to be poor, so comparing the relative binding free energies for sufficiently different systems is likely to produce unreliable results. Although the calculation of absolute binding free energies is notoriously difficult, methods in the field are improving all the time.²³¹ One particularly promising method, developed by Christopher Woods here at Bristol is the water swap method.²³² This could be applied to the systems described here in an attempt to calculate accurate binding free energies. A current technical barrier to doing this with the data from MD already generated in this project is that the water swap method is currently incompatible with the CHARMM force field, which was used in these simulations.

If an accurate method for calculating absolute or even relative binding free energies can be found for these systems, this could be employed in the aid of design and testing of

novel inhibitors for IDO and TDO. Even without accurate binding free energy calculations, improvements to docking methodologies could be made either with new MD simulations or by performing docking calculations on structures generated from the simulations described in this work.

References

- (1) Leach, A. R. *Molecular Modelling Principles and Applications*; second.; Pearson, 2001.
- (2) Newton, I. *Philosophiae Naturalis Principia Mathematica*; 1687.
- (3) Karplus, M. Development of Multiscale Models for Complex Chemical Systems: From H+H₂ to Biomolecules (Nobel Lecture). *Angew. Chem. Int. Ed.* **2014**, *53*, 9992–10005.
- (4) Habershon, S.; Manolopoulos, D. E.; Markland, T. E.; Miller, T. F. Ring-Polymer Molecular Dynamics: Quantum Effects in Chemical Dynamics from Classical Trajectories in an Extended Phase Space. *Annu. Rev. Phys. Chem.* **2013**, *64*, 387–413.
- (5) Althorpe, S. C.; Clary, D. C. Quantum Scattering Calculations on Chemical Reactions. *Annu. Rev. Phys. Chem.* **2003**, *54*, 493–529.
- (6) Habershon, S. Zero-Point Energy Effects in Anion Solvation Shells. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9154–9160.
- (7) Habershon, S.; Manolopoulos, D. E. Free Energy Calculations for a Flexible Water Model. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19714–19727.
- (8) Frank Jensen. *Introduction to Computational Chemistry*; 2nd ed.; Wiley, 2007.
- (9) Daan Frenkel; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press, 2001.
- (10) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; oxford science publications, 1989.
- (11) Watts, J. D.; Gauss, J.; Bartlett, R. J. Coupled-cluster Methods with Noniterative Triple Excitations for Restricted Open-shell Hartree–Fock and Other General Single Determinant Reference Functions. Energies and Analytical Gradients. *J. Chem. Phys.* **1993**, *98*, 8718–8733.
- (12) Ramabhadran, R. O.; Raghavachari, K. Extrapolation to the Gold-Standard in Quantum Chemistry: Computationally Efficient and Accurate CCSD(T) Energies for Large Molecules Using an Automated Thermochemical Hierarchy. *J. Chem. Theory Comput.* **2013**, *9*, 3986–3994.
- (13) Rauhut, G.; Knizia, G.; Werner, H.-J. Accurate Calculation of Vibrational Frequencies Using Explicitly Correlated Coupled-Cluster Theory. *J. Chem. Phys.* **2009**, *130*, 054105.
- (14) Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. Explicitly Correlated Electrons in Molecules. *Chem. Rev.* **2012**, *112*, 4–74.
- (15) Knizia, G.; Adler, T. B.; Werner, H.-J. Simplified CCSD(T)-F12 Methods: Theory and Benchmarks. *J. Chem. Phys.* **2009**, *130*, 054104.
- (16) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- (17) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.
- (18) Cramer, C. J.; Truhlar, D. G. Density Functional Theory for Transition Metals and Transition Metal Chemistry. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10757–10816.
- (19) Hirao, H. Which DFT Functional Performs Well in the Calculation of Methylcobalamin? Comparison of the B3LYP and BP86 Functionals and

- Evaluation of the Impact of Empirical Dispersion Correction. *J. Phys. Chem. A* **2011**, *115*, 9308–9313.
- (20) Siegbahn, P. E. M.; Blomberg, M. R. A.; Chen, S.-L. Significant van Der Waals Effects in Transition Metal Complexes. *J. Chem. Theory Comput.* **2010**, *6*, 2040–2044.
- (21) Zhao, Y.; Truhlar, D. G. Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States. *J. Phys. Chem. A* **2006**, *110*, 13126–13130.
- (22) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H–Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (23) Siegbahn, P. E. M.; Himo, F. Recent Developments of the Quantum Chemical Cluster Approach for Modeling Enzyme Reactions. *JBIC J. Biol. Inorg. Chem.* **2009**, *14*, 643–651.
- (24) Hine, N. D. M.; Haynes, P. D.; Mostofi, A. A.; Skylaris, C.-K.; Payne, M. C. Linear-Scaling Density-Functional Theory with Tens of Thousands of Atoms: Expanding the Scope and Scale of Calculations with ONETEP. *Comput. Phys. Commun.* **2009**, *180*, 1041–1053.
- (25) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (26) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Bioinforma.* **2006**, *65*, 712–725.
- (27) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380.
- (28) Jones, J. E. On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proc. R. Soc. Lond. Ser. A* **1924**, *106*, 463–477.
- (29) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, NA – NA.
- (30) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
- (31) Harvey, J. N. Spin-Forbidden CO Ligand Recombination in Myoglobin. *Faraday Discuss.* **2004**, *127*, 165–177.
- (32) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. *J. Phys. Chem. A* **2000**, *104*, 1720–1735.
- (33) Bucher, D.; Sandala, G. M.; Durbej, B.; Radom, L.; Smith, D. M. The Elusive 5'-Deoxyadenosyl Radical in Coenzyme-B12-Mediated Reactions. *J Am Chem Soc* **2012**, *134*, 1591–1599.

- (34) Evans, M. G.; Polanyi, M. Inertia and Driving Force of Chemical Reactions. *Trans. Faraday Soc.* **1938**, *34*, 11–24.
- (35) Sato, S. A New Method of Drawing the Potential Energy Surface. *Bull. Chem. Soc. Jpn.* **1955**, *28*, 450–453.
- (36) Wong, K. F.; Sonnenberg, J. L.; Paesani, F.; Yamamoto, T.; Vanicek, J.; Zhang, W.; Schlegel, H. B.; Case, D. A.; Cheatham, T. E.; Miller, W. H.; Voth, G. A. Proton Transfer Studied Using a Combined Ab Initio Reactive Potential Energy Surface with Quantum Path Integral Methodology. *J. Chem. Theory Comput.* **2010**, *6*, 2566–2580.
- (37) Kamerlin, S. C. L.; Warshel, A. The EVB as a Quantitative Tool for Formulating Simulations and Analyzing Biological and Chemical Reactions. *Faraday Discuss.* **2010**, *145*, 71.
- (38) Schlegel, H. B.; Sonnenberg, J. L. Empirical Valence-Bond Models for Reactive Potential Energy Surfaces Using Distributed Gaussians. *J. Chem. Theory Comput.* **2006**, *2*, 905–911.
- (39) Van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.
- (40) Glowacki, D. R.; Orr-Ewing, A. J.; Harvey, J. N. Product Energy Deposition of CN + Alkane H Abstraction Reactions in Gas and Solution Phases. *J. Chem. Phys.* **2011**, *134*, 214508.
- (41) Warshel, A.; Weiss, R. M. An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.
- (42) Schmitt, U. W.; Voth, G. A. Multistate Empirical Valence Bond Model for Proton Transport in Water. *J. Phys. Chem. B* **1998**, *102*, 5547–5551.
- (43) I Zúñiga, P. E. Scaling of the Time-Dependent Diffusion Coefficient by Molecular Dynamics Simulation. *Phys. Rev. Lett.* **1993**, *71*, 3665–3668.
- (44) Csermely, P.; Palotai, R.; Nussinov, R. Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. *Trends Biochem. Sci.* **2010**, *35*, 539–546.
- (45) Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 107–115.
- (46) Hele, T. J. H.; Althorpe, S. C. Derivation of a True ($t \rightarrow 0+$) Quantum Transition-State Theory. I. Uniqueness and Equivalence to Ring-Polymer Molecular Dynamics Transition-State-Theory. *J. Chem. Phys.* **2013**, *138*, 084108.
- (47) Hele, T. J. H.; Althorpe, S. C. On the Uniqueness of $T \rightarrow 0+$ Quantum Transition-State Theory. *J. Chem. Phys.* **2013**, *139*, 084116.
- (48) Schlegel, H. B. Exploring Potential Energy Surfaces for Chemical Reactions: An Overview of Some Practical Methods. *J. Comput. Chem.* **2003**, *24*, 1514–1527.
- (49) Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. On Possible Pitfalls in Ab Initio Quantum Mechanics/Molecular Mechanics Minimization Approaches for Studies of Enzymatic Reactions. *J. Phys. Chem. B* **2005**, *109*, 15645–15650.
- (50) Lonsdale, R. *Cytochrome P450 Reactivity and Specificity from QM/MM Modelling*. A PhD Thesis; 2009.
- (51) Cooper, A. M.; Kästner, J. Averaging Techniques for Reaction Barriers in QM/MM Simulations. *ChemPhysChem* **2014**, n/a – n/a.
- (52) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Compound I Reactivity Defines Alkene Oxidation Selectivity in Cytochrome P450cam. *J. Phys. Chem. B* **2010**, *114*, 1156–1162.
- (53) Ken Dill; Bromberg, S. *Molecular Driving Forces*; second.; Garland Science, 2011.

- (54) Glowacki, D. R.; Paci, E.; Shalashilin, D. V. Boxed Molecular Dynamics: A Simple and General Technique for Accelerating Rare Event Kinetics and Mapping Free Energy in Large Molecular Systems. *J. Phys. Chem. B* **2009**, *113*, 16603–16611.
- (55) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71*, 126601.
- (56) Park, S.; Schulten, K. Calculating Potentials of Mean Force from Steered Molecular Dynamics Simulations. *J. Chem. Phys.* **2004**, *120*, 5946–5961.
- (57) Kästner, J.; Thiel, W. Bridging the Gap between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: “Umbrella Integration.” *J. Chem. Phys.* **2005**, *123*, 144104.
- (58) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (59) Marsh, E. N. G.; Patterson, D. P.; Li, L. Adenosyl Radical: Reagent and Catalyst in Enzyme Reactions. *ChemBioChem* **2010**, *11*, 604–621.
- (60) Buckel, W.; Golding, B. T. Radical Enzymes in Anaerobes*. *Annu. Rev. Microbiol.* **2006**, *60*, 27–49.
- (61) Hay, B. P.; Finke, R. G. Thermolysis of the Cobalt-Carbon Bond of Adenosylcobalamin. 2. Products, Kinetics, and Cobalt-Carbon Bond Dissociation Energy in Aqueous Solution. *J Am Chem Soc* **1986**, *108*, 4820–4829.
- (62) Marsh, E. N. G.; Meléndez, G. D. R. Adenosylcobalamin Enzymes: Theory and Experiment Begin to Converge. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2012**, *1824*, 1154–1164.
- (63) Li, X.; Chung, L. W.; Paneth, P.; Morokuma, K. <title> DFT and ONIOM(DFT:MM) Studies on Co–C Bond Cleavage and Hydrogen Transfer in B₁₂-Dependent Methylmalonyl-CoA Mutase. Stepwise or Concerted Mechanism? </title>. *J. Am. Chem. Soc.* **2009**, *131*, 5115–5125.
- (64) Sharma, P. K.; Chu, Z. T.; Olsson, M. H. M.; Warshel, A. A New Paradigm for Electrostatic Catalysis of Radical Reactions in Vitamin B12 Enzymes. *Proc. Natl. Acad. Sci.* **2007**, *104*, 9661–9666.
- (65) Cheng, M.-C.; Marsh, E. N. G. Evidence for Coupled Motion and Hydrogen Tunneling of the Reaction Catalyzed by Glutamate Mutase†. *Biochemistry (Mosc.)* **2006**, *46*, 883–889.
- (66) Banerjee, R.; Dybala-Defratyka, A.; Paneth, P. Quantum Catalysis in B12-Dependent Methylmalonyl-CoA Mutase: Experimental and Computational Insights. *Philos. Trans. R. Soc. B Biol. Sci.* **2006**, *361*, 1333–1339.
- (67) Chowdhury, S.; Banerjee, R. Evidence for Quantum Mechanical Tunneling in the Coupled Cobalt–Carbon Bond Homolysis–Substrate Radical Generation Reaction Catalyzed by Methylmalonyl-CoA Mutase. *J. Am. Chem. Soc.* **2000**, *122*, 5417–5418.
- (68) Xia, L.; Ballou, D. P.; Marsh, E. N. G. Role of Arg100 in the Active Site of Adenosylcobalamin-Dependent Glutamate Mutase†. *Biochemistry (Mosc.)* **2004**, *43*, 3238–3245.
- (69) Marsh, E. N. G. Insights into the Mechanisms of Adenosylcobalamin (coenzyme B12)-Dependent Enzymes from Rapid Chemical Quench Experiments. *Biochem. Soc. Trans.* **2009**, *37*, 336.
- (70) Sension, R. J.; Harris, D. A.; Stickrath, A.; Cole, A. G.; Fox, C. C.; Marsh, E. N. G. Time-Resolved Measurements of the Photolysis and Recombination of

- Adenosylcobalamin Bound to Glutamate Mutase. *J Phys Chem B* **2005**, *109*, 18146–18152.
- (71) Sandala, G. M.; Smith, D. M.; Marsh, E. N. G.; Radom, L. Toward an Improved Understanding of the Glutamate Mutase System. *J Am Chem Soc* **2007**, *129*, 1623–1633.
- (72) Chih, H.-W.; Marsh, E. N. G. Mechanism of Glutamate Mutase: Identification and Kinetic Competence of Acrylate and Glycyl Radical as Intermediates in the Rearrangement of Glutamate to Methylaspartate. *J Am Chem Soc* **2000**, *122*, 10732–10733.
- (73) Madhavapeddi, P.; Marsh, E. N. G. The Role of the Active Site Glutamate in the Rearrangement of Glutamate to 3-Methylaspartate Catalyzed by Adenosylcobalamin-Dependent Glutamate Mutase. *Chem. Biol.* **2001**, *8*, 1143–1149.
- (74) Marsh, E. N. G.; Ballou, D. P. Coupling of Cobalt–Carbon Bond Homolysis and Hydrogen Atom Abstraction in Adenosylcobalamin-Dependent Glutamate Mutase†. *Biochemistry (Mosc.)* **1998**, *37*, 11864–11872.
- (75) Sandala, G. M.; Smith, D. M.; Radom, L. Modeling the Reactions Catalyzed by Coenzyme B12-Dependent Enzymes. *Acc. Chem. Res.* **2010**, *43*, 642–651.
- (76) Rommel, J. B.; Kästner, J. The Fragmentation–Recombination Mechanism of the Enzyme Glutamate Mutase Studied by QM/MM Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 10195–10203.
- (77) Jensen, K. P.; Ryde, U. Cobalamins Uncovered by Modern Electronic Structure Calculations. *Coord. Chem. Rev.* **2009**, *253*, 769–778.
- (78) Jensen, K. P.; Ryde, U. The Axial N-Base Has Minor Influence on Co–C Bond Cleavage in Cobalamins. *J. Mol. Struct. THEOCHEM* **2002**, *585*, 239–255.
- (79) Jensen, K. P.; Ryde, U. Theoretical Prediction of the Co–C Bond Strength in Cobalamins. *J. Phys. Chem. A* **2003**, *107*, 7539–7545.
- (80) Siegbahn, P. E. M.; Blomberg, M. R. A.; Chen, S.-L. Significant van Der Waals Effects in Transition Metal Complexes. *J Chem Theory Comput* **2010**, *6*, 2040–2044.
- (81) Kepp, K. P. Co–C Dissociation of Adenosylcobalamin (Coenzyme B12): Role of Dispersion, Induction Effects, Solvent Polarity, and Relativistic and Thermal Corrections. *J. Phys. Chem. A* **2014**, *118*, 7104–7117.
- (82) Kozlowski, P. M.; Kamachi, T.; Toraya, T.; Yoshizawa, K. Does Cob(II)alamin Act as a Conductor in Coenzyme B12 Dependent Mutases? *Angew. Chem. Int. Ed.* **2007**, *46*, 980–983.
- (83) Jensen, K. P.; Ryde, U. <title> How the Co–C Bond Is Cleaved in Coenzyme B₁₂ Enzymes: A Theoretical Study </title>. *J. Am. Chem. Soc.* **2005**, *127*, 9117–9128.
- (84) Kwiecien, R. A.; Khavrutskii, I. V.; Musaev, D. G.; Morokuma, K.; Banerjee, R.; Paneth, P. Computational Insights into the Mechanism of Radical Generation in B12-Dependent Methylmalonyl-CoA Mutase. *J. Am. Chem. Soc.* **2006**, *128*, 1287–1292.
- (85) Pang, J.; Li, X.; Morokuma, K.; Scrutton, N. S.; Sutcliffe, M. J. Large-Scale Domain Conformational Change Is Coupled to the Activation of the Co–C Bond in the B12-Dependent Enzyme Ornithine 4,5-Aminomutase: A Computational Study. *J. Am. Chem. Soc.* **2012**, *134*, 2367–2377.
- (86) ClaeysSENS, F.; Ranaghan, K. E.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Multiple High-Level QM/MM Reaction Paths Demonstrate Transition-State Stabilization in Chorismate Mutase: Correlation of Barrier Height with Transition-State Stabilization. *Chem. Commun.* **2005**, 5068–5070.

- (87) Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. P450 Enzymes: Their Structure, Reactivity, and Selectivity—Modeled by QM/MM Calculations. *Chem. Rev.* **2010**, *110*, 949–1017.
- (88) Van der Kamp, M. W.; Chaudret, R.; Mulholland, A. J. QM/MM Modelling of Ketosteroid Isomerase Reactivity Indicates That Active Site Closure Is Integral to Catalysis. *FEBS J.* **2013**, *280*, 3120–3131.
- (89) Warshel, A. COMPUTER SIMULATIONS OF ENZYME CATALYSIS: Methods, Progress, and Insights. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.
- (90) Valero, R.; Song, L.; Gao, J.; Truhlar, D. G. Perspective on Diabatic Models of Chemical Reactivity as Illustrated by the Gas-Phase SN2 Reaction of Acetate Ion with 1,2-Dichloroethane. *J. Chem. Theory Comput.* **2009**, *5*, 1–22.
- (91) Pereira, C. S.; Kony, D.; Baron, R.; Müller, M.; van Gunsteren, W. F.; Hünenberger, P. H. Conformational and Dynamical Properties of Disaccharides in Water: A Molecular Dynamics Study. *Biophys. J.* **2006**, *90*, 4337–4344.
- (92) Wolthers, K. R.; Levy, C.; Scrutton, N. S.; Leys, D. Large-Scale Domain Dynamics and Adenosylcobalamin Reorientation Orchestrate Radical Catalysis in Ornithine 4,5-Aminomutase. *J. Biol. Chem.* **2010**, *285*, 13942–13950.
- (93) Brunk, E.; Kellett, W. F.; Richards, N. G. J.; Rothlisberger, U. A Mechanochemical Switch to Control Radical Intermediates. *Biochemistry (Mosc.)* **2014**, *53*, 3830–3838.
- (94) Garr, C. D.; Finke, R. G. Adocobalamin (AdoCbl or Coenzyme B12) Cobalt-Carbon Bond Homolysis Radical-Cage Effects: Product, Kinetic, Mechanistic, and Cage Efficiency Factor (Fc) Studies, plus the Possibility That Coenzyme B12-Dependent Enzymes Function as “Ultimate Radical Cages” and “Ultimate Radical Traps.” *Inorg. Chem.* **1993**, *32*, 4414–4421.
- (95) Martin, B. D.; Finke, R. G. Methylcobalamin’s Full- vs. Half-Strength Cobalt-Carbon Sigma Bonds and Bond Dissociation Enthalpies: A $>10^{15}$ Co-CH₃ Homolysis Rate Enhancement Following One-Antibonding-Electron Reduction of Methylcobalamin. *J. Am. Chem. Soc.* **1992**, *114*, 585–592.
- (96) Reiher, M.; Salomon, O.; Hess, B. A. Reparameterization of Hybrid Functionals Based on Energy Differences of States of Different Multiplicity. *Theor. Chem. Acc.* **2001**, *107*, 48–55.
- (97) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (98) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (99) Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

- (100) Hehre, W.; Ditchfie.r; Pople, J. Self-Consistent Molecular-Orbital Methods .12. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular-Orbital Studies of Organic-Molecules. *J. Chem. Phys.* **1972**, *56*, 2257 – &.
- (101) Hay, P. J.; Wadt, W. R. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985**, *82*, 270–283.
- (102) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. *MOLPRO, Version 2012.1, a Package of Ab Initio Programs*; molpro, 2012.
- (103) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. Molpro: A General-purpose Quantum Chemistry Program Package. *WIREs Comput Mol Sci* **2012**, *2*, 242–253.
- (104) Jr, T. H. D. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (105) Balabanov, N. B.; Peterson, K. A. Systematically Convergent Basis Sets for Transition Metals. I. All-Electron Correlation Consistent Basis Sets for the 3d Elements Sc–Zn. *J. Chem. Phys.* **2005**, *123*, 064107.
- (106) Hughes, T. F.; Harvey, J. N.; Friesner, R. A. A B3LYP-DBLOC Empirical Correction Scheme for Ligand Removal Enthalpies of Transition Metal Complexes: Parameterization against Experimental and CCSD(T)-F12 Heats of Formation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7724–7738.
- (107) Peng, D.; Hirao, K. An Arbitrary Order Douglas–Kroll Method with Polynomial Cost. *J. Chem. Phys.* **2009**, *130*, 044102.
- (108) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (109) Marques, H. M.; Ngoma, B.; Egan, T. J.; Brown, K. L. Parameters for the Amber Force Field for the Molecular Mechanics Modeling of the Cobalt Corrinoids. *J. Mol. Struct.* **2001**, *561*, 71–91.
- (110) Pavelites, J. J.; Gao, J.; Bash, P. A.; Mackerell, A. D. A Molecular Mechanics Force Field for NAD⁺ NADH, and the Pyrophosphate Groups of Nucleotides. *J. Comput. Chem.* **1997**, *18*, 221–239.
- (111) Walker, R. C.; de Souza, M. M.; Mercer, I. P.; Gould, I. R.; Klug, D. R. Large and Fast Relaxations inside a Protein: Calculation and Measurement of Reorganization Energies in Alcohol Dehydrogenase. *J. Phys. Chem. B* **2002**, *106*, 11658–11665.
- (112) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y. R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.* **2011**, *39*, W511–W517.
- (113) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *Int. J. Quantum Chem.* **2013**, *113*, 2110–2142.

- (114) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
- (115) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell—a Modular Software Package for QM/MM Simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 101–110.
- (116) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994**, *100*, 9025–9031.
- (117) Gruber, K.; Reitzer, R.; Kratky, C. Radical Shuttling in a Protein: Ribose Pseudorotation Controls Alkyl-Radical Transfer in the Coenzyme B12 Dependent Enzyme Glutamate Mutase. *Angew. Chem. Int. Ed.* **2001**, *40*, 3377–3380.
- (118) Román-Meléndez, G. D.; von Glehn, P.; Harvey, J. N.; Mulholland, A. J.; Marsh, E. N. G. Role of Active Site Residues in Promoting Cobalt–Carbon Bond Homolysis in Adenosylcobalamin-Dependent Mutases Revealed through Experiment and Computation. *Biochemistry (Mosc.)* **2014**, *53*, 169–177.
- (119) Sharma, P. K.; Chu, Z. T.; Olsson, M. H. M.; Warshel, A. A New Paradigm for Electrostatic Catalysis of Radical Reactions in Vitamin B-12 Enzymes RID E-4501-2011. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9661–9666.
- (120) Mancia, F.; Evans, P. R. Conformational Changes on Substrate Binding to Methylmalonyl CoA Mutase and New Insights into the Free Radical Mechanism. *Struct. Lond. Engl.* **1998**, *6*, 711–720.
- (121) Hodgkin, D. C. The X-Ray Analysis of Complicated Molecules. *Science* **1965**, *150*, 979–988.
- (122) Jensen, K. P.; Ryde, U. How the Co–C Bond Is Cleaved in Coenzyme B12 Enzymes: A Theoretical Study. *J. Am. Chem. Soc.* **2005**, *127*, 9117–9128.
- (123) Sharma, P. K.; Chu, Z. T.; Olsson, M. H. M.; Warshel, A. A New Paradigm for Electrostatic Catalysis of Radical Reactions in Vitamin B12 Enzymes. *Proc. Natl. Acad. Sci.* **2007**, *104*, 9661–9666.
- (124) Rostkowski, M.; Olsson, M. H.; Søndergaard, C. R.; Jensen, J. H. Graphical Analysis of pH-Dependent Properties of Proteins Predicted Using PROPKA. *BMC Struct. Biol.* **2011**, *11*, 6.
- (125) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity : All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2009**, *66*, 12–21.
- (126) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (127) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38, 27–28.
- (128) Altona, C.; Sundaralingam, M. Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. New Description Using the Concept of Pseudorotation. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.
- (129) Roe, D. R.; Cheatham, T. E. PTraj and CPPPTraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
- (130) Rétey, J. Enzymic Reaction Selectivity by Negative Catalysis or How Do Enzymes Deal with Highly Reactive Intermediates? *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 355–361.

- (131) Li, X.; Chung, L. W.; Paneth, P.; Morokuma, K. DFT and ONIOM(DFT:MM) Studies on Co–C Bond Cleavage and Hydrogen Transfer in B12-Dependent Methylmalonyl-CoA Mutase. Stepwise or Concerted Mechanism? *J. Am. Chem. Soc.* **2009**, *131*, 5115–5125.
- (132) Marsh, E. N. G.; Patterson, D. P.; Li, L. Adenosyl Radical: Reagent and Catalyst in Enzyme Reactions. *ChemBioChem* **2010**, *11*, 604–621.
- (133) Cheng, M.-C.; Marsh, E. N. G. Pre-Steady-State Measurement of Intrinsic Secondary Tritium Isotope Effects Associated with the Homolysis of Adenosylcobalamin and the Formation of 5'-Deoxyadenosine in Glutamate Mutase†. *Biochemistry (Mosc.)* **2004**, *43*, 2155–2158.
- (134) Sandala, G. M.; Smith, D. M.; Marsh, E. N. G.; Radom, L. Toward an Improved Understanding of the Glutamate Mutase System. *J Am Chem Soc* **2007**, *129*, 1623–1633.
- (135) Sandala, G. M.; Smith, D. M.; Radom, L. Modeling the Reactions Catalyzed by Coenzyme B12-Dependent Enzymes. *Acc. Chem. Res.* **2010**, *43*, 642–651.
- (136) Chih, H. W.; Roymoulik, I.; Huhta, M. S.; Madhavapeddi, P.; Marsh, E. N. G. Adenosylcobalamin-Dependent Glutamate Mutase: Pre-Steady-State Kinetic Methods for Investigating Reaction Mechanism. *Enzyme Kinet. Mech. Pt F Detect. Charact. Enzyme React. Intermed.* **2002**, *354*, 380–399.
- (137) Jensen, K. P.; Ryde, U. How the Co–C Bond Is Cleaved in Coenzyme B12 Enzymes: A Theoretical Study. *J. Am. Chem. Soc.* **2005**, *127*, 9117–9128.
- (138) Ponder, J. W. *TINKER - Software Tools for Molecular Design Version 6.1*; 2012.
- (139) Harvey, J. N. Spin-Forbidden CO Ligand Recombination in Myoglobin. *Faraday Discuss.* **2004**, *127*, 165–177.
- (140) Iii, C. L. B.; Karplus, M. Deformable Stochastic Boundaries in Molecular Dynamics. *J. Chem. Phys.* **1983**, *79*, 6312–6325.
- (141) RANAGHAN, K. E.; RIDDER, L.; SZEFCZYK, B.; SOKALSKI, W. A.; HERMANN, J. C.; MULHOLLAND, A. J. Insights into Enzyme Catalysis from QM/MM Modelling: Transition State Stabilization in Chorismate Mutase. *Mol. Phys.* **2003**, *101*, 2695–2714.
- (142) Nakagawa, S.; Yu, H.-A.; Karplus, M.; Urneyama, H. Active Site Dynamics of Acyl-Chymotrypsin. *Proteins Struct. Funct. Bioinforma.* **1993**, *16*, 172–194.
- (143) Chen, H.-P.; Marsh, E. N. G. Adenosylcobalamin-Dependent Glutamate Mutase: Examination of Substrate and Coenzyme Binding in an Engineered Fusion Protein Possessing Simplified Subunit Structure and Kinetic Properties†. *Biochemistry (Mosc.)* **1997**, *36*, 14939–14945.
- (144) Cheng, M.-C.; Marsh, E. N. G. Evidence for Coupled Motion and Hydrogen Tunneling of the Reaction Catalyzed by Glutamate Mutase†. *Biochemistry (Mosc.)* **2006**, *46*, 883–889.
- (145) Bothe, H.; Darley, D. J.; Albracht, S. P.; Gerfen, G. J.; Golding, B. T.; Buckel, W. Identification of the 4-Glutamyl Radical as an Intermediate in the Carbon Skeleton Rearrangement Catalyzed by Coenzyme B12-Dependent Glutamate Mutase from Clostridium Cochlearium. *Biochemistry (Mosc.)* **1998**, *37*, 4105–4113.
- (146) Yoon, M.; Song, H.; Håkansson, K.; Marsh, E. N. G. Hydrogen Tunneling in Adenosylcobalamin-Dependent Glutamate Mutase: Evidence from Intrinsic Kinetic Isotope Effects Measured by Intramolecular Competition. *Biochemistry (Mosc.)* **2010**, *49*, 3168–3173.
- (147) Doll, K.; Bender, B.; Finke, R. The First Experimental Test of the Hypothesis That Enzymes Have Evolved to Enhance Hydrogen Tunneling. *J. Am. Chem. Soc.* **2003**, *125*, 10877–10884.

- (148) Rommel, J. B.; Liu, Y.; Werner, H.-J.; Kästner, J. Role of Tunneling in the Enzyme Glutamate Mutase. *J. Phys. Chem. B* **2012**, *116*, 13682–13689.
- (149) Dybala-Defratyka, A.; Paneth, P.; Banerjee, R.; Truhlar, D. G. Coupling of Hydrogenic Tunneling to Active-Site Motion in the Hydrogen Radical Transfer Catalyzed by a Coenzyme B12-Dependent Mutase. *Proc. Natl. Acad. Sci.* **2007**, *104*, 10774–10779.
- (150) Sandala, G. M.; Smith, D. M.; Coote, M. L.; Golding, B. T.; Radom, L. Insights into the Hydrogen-Abstraction Reactions of Diol Dehydratase: Relevance to the Catalytic Mechanism and Suicide Inactivation. *J. Am. Chem. Soc.* **2006**, *128*, 3433–3444.
- (151) Craig, I. R.; Manolopoulos, D. E. Chemical Reaction Rates from Ring Polymer Molecular Dynamics. *J. Chem. Phys.* **2005**, *122*, 084106.
- (152) Young, V. R. Adult Amino Acid Requirements: The Case for a Major Revision in Current Recommendations. *J. Nutr.* **1994**, *124*, 1517S – 1523S.
- (153) Peters, J. C. Tryptophan Nutrition and Metabolism: An Overview. *Adv. Exp. Med. Biol.* **1991**, *294*, 345–358.
- (154) Pilote, L.; Larrieu, P.; Stroobant, V.; Colau, D.; Dolušić, E.; Frédérick, R.; Plaen, E. D.; Uyttenhove, C.; Wouters, J.; Masereel, B.; Eynde, B. J. V. den. Reversal of Tumoral Immune Resistance by Inhibition of Tryptophan 2,3-Dioxygenase. *Proc. Natl. Acad. Sci.* **2012**, *109*, 2497–2502.
- (155) Munn, D. H.; Zhou, M.; Attwood, J. T.; Bondarev, I.; Conway, S. J.; Marshall, B.; Brown, C.; Mellor, A. L. Prevention of Allogeneic Fetal Rejection by Tryptophan Catabolism. *Science* **1998**, *281*, 1191–1193.
- (156) Metz, R.; Smith, C.; DuHadaway, J. B.; Chandler, P.; Baban, B.; Merlo, L. M. F.; Pigott, E.; Keough, M. P.; Rust, S.; Mellor, A. L.; Mandik-Nayak, L.; Muller, A. J.; Prendergast, G. C. IDO2 Is Critical for IDO1-Mediated T-Cell Regulation and Exerts a Non-Redundant Function in Inflammation. *Int. Immunol.* **2014**, dxt073.
- (157) Platten, M.; Wick, W.; Eynde, B. J. V. den. Tryptophan Catabolism in Cancer: Beyond IDO and Tryptophan Depletion. *Cancer Res.* **2012**, *72*, 5435–5440.
- (158) Novitskiy, S. V.; Moses, H. L. Turn Off the IDO: Will Clinical Trials Be Successful? *Cancer Discov.* **2012**, *2*, 673–675.
- (159) Opitz, C. A.; Litzenburger, U. M.; Opitz, U.; Sahm, F.; Ochs, K.; Lutz, C.; Wick, W.; Platten, M. The Indoleamine-2,3-Dioxygenase (IDO) Inhibitor 1-Methyl-D-Tryptophan Upregulates IDO1 in Human Cancer Cells. *PLoS ONE* **2011**, *6*, e19823.
- (160) Sugimoto, H.; Oda, S.; Otsuki, T.; Hino, T.; Yoshida, T.; Shiro, Y. Crystal Structure of Human Indoleamine 2,3-Dioxygenase: Catalytic Mechanism of O₂ Incorporation by a Heme-Containing Dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 2611–2616.
- (161) Forouhar, F.; Anderson, J. L. R.; Mowat, C. G.; Vorobiev, S. M.; Hussain, A.; Abashidze, M.; Bruckmann, C.; Thackray, S. J.; Seetharaman, J.; Tucker, T.; Xiao, R.; Ma, L.-C.; Zhao, L.; Acton, T. B.; Montelione, G. T.; Chapman, S. K.; Tong, L. Molecular Insights into Substrate Recognition and Catalysis by Tryptophan 2,3-Dioxygenase. *Proc. Natl. Acad. Sci.* **2007**, *104*, 473–478.
- (162) Sugimoto, H.; Oda, S.; Otsuki, T.; Hino, T.; Yoshida, T.; Shiro, Y. Crystal Structure of Human Indoleamine 2,3-Dioxygenase: Catalytic Mechanism of O₂ Incorporation by a Heme-Containing Dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 2611–2616.
- (163) Zhang, Y.; Kang, S. A.; Mukherjee, T.; Bale, S.; Crane, B. R.; Begley, T. P.; Ealick, S. E. Crystal Structure and Mechanism of Tryptophan 2,3-Dioxygenase, a

- Heme Enzyme Involved in Tryptophan Catabolism and in Quinolinate Biosynthesis(,). *Biochemistry (Mosc.)* **2007**, *46*, 145–155.
- (164) Terentis, A. C.; Thomas, S. R.; Takikawa, O.; Littlejohn, T. K.; Truscott, R. J. W.; Armstrong, R. S.; Yeh, S.-R.; Stocker, R. The Heme Environment of Recombinant Human Indoleamine 2,3-Dioxygenase. *J. Biol. Chem.* **2002**, *277*, 15788–15794.
- (165) Davydov, R. M.; Chauhan, N.; Thackray, S. J.; Anderson, J. L. ; Papadopoulou, N. D.; Mowat, C. G.; Chapman, S. K.; Raven, E. L.; Hoffman, B. M. Probing the Ternary Complexes of Indoleamine and Tryptophan 2, 3-Dioxygenases by Cryoreduction EPR and ENDOR Spectroscopy. *J. Am. Chem. Soc.* **2010**, *132*, 5494–5500.
- (166) Littlejohn, T. K.; Takikawa, O.; Truscott, R. J. W.; Walker, M. J. Asp274 and His346 Are Essential for Heme Binding and Catalytic Function of Human Indoleamine 2,3-Dioxygenase. *J. Biol. Chem.* **2003**, *278*, 29525–29531.
- (167) *Amino Acids and the Asymmetry of Life - Caught in the Act of Formation.*
- (168) Chauhan, N.; Thackray, S. J.; Rafice, S. A.; Eaton, G.; Lee, M.; Efimov, I.; Basran, J.; Jenkins, P. R.; Mowat, C. G.; Chapman, S. K.; Raven, E. L. Reassessment of the Reaction Mechanism in the Heme Dioxygenases. *J. Am. Chem. Soc.* **2009**, *131*, 4186+.
- (169) Basran, J.; Rafice, S. A.; Chauhan, N.; Efimov, I.; Cheesman, M. R.; Ghamsari, L.; Raven, E. L. A Kinetic, Spectroscopic, and Redox Study of Human Tryptophan 2,3-Dioxygenase. *Biochemistry (Mosc.)* **2008**, *47*, 4752–4760.
- (170) Thackray, S. J.; Bruckmann, C.; Anderson, J. L. R.; Campbell, L. P.; Xiao, R.; Zhao, L.; Mowat, C. G.; Forouhar, F.; Tong, L.; Chapman, S. K. Histidine 55 of Tryptophan 2,3-Dioxygenase Is Not an Active Site Base but Regulates Catalysis by Controlling Substrate Binding. *Biochemistry (Mosc.)* **2008**, *47*, 10677–10684.
- (171) Sono, M.; Roach, M. P.; Coulter, E. D.; Dawson, J. H. Heme-Containing Oxygenases. *Chem. Rev.* **1996**, *96*, 2841–2888.
- (172) Cady, S. G.; Sono, M. 1-Methyl-DL-Tryptophan, Beta-(3-Benzofuranyl)-DL-Alanine (the Oxygen Analog of Tryptophan), and Beta-[3-Benzo(b)thienyl]-DL-Alanine (the Sulfur Analog of Tryptophan) Are Competitive Inhibitors for Indoleamine 2,3-Dioxygenase. *Arch. Biochem. Biophys.* **1991**, *291*, 326–333.
- (173) Chung, L. W.; Li, X.; Sugimoto, H.; Shiro, Y.; Morokuma, K. Density Functional Theory Study on a Missing Piece in Understanding of Heme Chemistry: The Reaction Mechanism for Indoleamine 2,3-Dioxygenase and Tryptophan 2,3-Dioxygenase. *J. Am. Chem. Soc.* **2008**, *130*, 12299–12309.
- (174) Lewis-Ballester, A.; Batabyal, D.; Egawa, T.; Lu, C.; Lin, Y.; Marti, M. A.; Capece, L.; Estrin, D. A.; Yeh, S.-R. Evidence for a Ferryl Intermediate in a Heme-Based Dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 17371–17376.
- (175) Capece, L.; Lewis-Ballester, A.; Yeh, S.-R.; Estrin, D. A.; Marti, M. A. Complete Reaction Mechanism of Indoleamine 2,3-Dioxygenase as Revealed by QM/MM Simulations. *J. Phys. Chem. B* **2012**, *116*, 1401–1413.
- (176) Guallar, V.; Wallrapp, F. H. QM/MM Methods: Looking inside Heme Proteins Biochemistry. *Biophys. Chem.* **2010**, *149*, 1–11.
- (177) Chung, L. W.; Li, X.; Sugimoto, H.; Shiro, Y.; Morokuma, K. ONIOM Study on a Missing Piece in Our Understanding of Heme Chemistry: Bacterial Tryptophan 2,3-Dioxygenase with Dual Oxidants. *J. Am. Chem. Soc.* **2010**, *132*, 11993–12005.
- (178) Chauhan, N.; Basran, J.; Efimov, I.; Svistunenko, D. A.; Seward, H. E.; Moody, P. C. E.; Raven, E. L. The Role of Serine 167 in Human Indoleamine 2,3-

- Dioxygenase: A Comparison with Tryptophan 2,3-Dioxygenase. *Biochemistry (Mosc.)* **2008**, *47*, 4761–4769.
- (179) Röhrlig, U. F.; Awad, L.; Grosdidier, A.; Larrieu, P.; Stroobant, V.; Colau, D.; Cerundolo, V.; Simpson, A. J. G.; Vogel, P.; Van den Eynde, B. J.; Zoete, V.; Michielin, O. Rational Design of Indoleamine 2,3-Dioxygenase Inhibitors. *J. Med. Chem.* **2010**, *53*, 1172–1189.
- (180) Dawood, S.; Zarina, S.; Bano, S. Docking Studies of Antidepressants against Single Crystal Structure of Tryptophan 2, 3-Dioxygenase Using Molegro Virtual Docker Software. *Pak. J. Pharm. Sci.* **2014**, *27*, 1529–1539.
- (181) John, S.; Thangapandian, S.; Sakkiah, S.; Lee, K. W. Identification of Potent Virtual Leads to Design Novel Indoleamine 2,3-Dioxygenase Inhibitors: Pharmacophore Modeling and Molecular Docking Studies. *Eur. J. Med. Chem.* **2010**, *45*, 4004–4012.
- (182) Macchiarulo, A.; Nuti, R.; Bellocchi, D.; Camaioni, E.; Pellicciari, R. Molecular Docking and Spatial Coarse Graining Simulations as Tools to Investigate Substrate Recognition, Enhancer Binding and Conformational Transitions in Indoleamine-2,3-Dioxygenase (IDO). *Biochim. Biophys. ACTA-PROTEINS PROTEOMICS* **2007**, *1774*, 1058–1068.
- (183) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining Docking and Molecular Dynamic Simulations in Drug Design. *Med. Res. Rev.* **2006**, *26*, 531–568.
- (184) Forouhar, F.; Anderson, J. L. R.; Mowat, C. G.; Vorobiev, S. M.; Hussain, A.; Abashidze, M.; Bruckmann, C.; Thackray, S. J.; Seetharaman, J.; Tucker, T.; Xiao, R.; Ma, L.-C.; Zhao, L.; Acton, T. B.; Montelione, G. T.; Chapman, S. K.; Tong, L. Molecular Insights into Substrate Recognition and Catalysis by Tryptophan 2,3-Dioxygenase. *Proc. Natl. Acad. Sci.* **2007**, *104*, 473–478.
- (185) (11) Molecular insights into substrate recognition and catalysis by tryptophan 2,3-dioxygenase — PNAS
<http://www.pnas.org/content/104/2/473.full.pdf+html?sid=8924cb69-b50b-4c91-a195-e4e98afb144e> (accessed Nov 4, 2009).
- (186) Quillin, M. L.; Arduini, R. M.; Olson, J. S.; Phillips Jr, G. N. High-Resolution Crystal Structures of Distal Histidine Mutants of Sperm Whale Myoglobin. *J. Mol. Biol.* **1993**, *234*, 140–155.
- (187) Phillips, S. E. Structure and Refinement of Oxymyoglobin at 1.6 Å Resolution. *J. Mol. Biol.* **1980**, *142*, 531–554.
- (188) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-Pdb Viewer: An Environment for Comparative Protein Modeling. *ELECTROPHORESIS* **1997**, *18*, 2714–2723.
- (189) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
- (190) Von Glehn, P. *Investigation into the Dynamic Structure of a Haem Dioxygenase Enzyme with Molecular Dynamics Simulations. A Final Year Report Submitted in Partial Fulfilment of the Requirements for an BSc in Chemistry at the University of Bristol.*
- (191) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom

- Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins[†]. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (192) Strickland, N. *QM and QM/MM Studies of Ligand Binding to Myoglobin and Haemoglobin*. A PhD Thesis Submitted to the University of Bristol; 2006.
- (193) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (194) Van der Kamp, M. W. *Modelling Reactions and Dynamics of Claisen Enzymes*. A PhD Thesis Submitted to the University of Bristol; 2008.
- (195) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (196) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
- (197) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.
- (198) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.
- (199) Massova, I.; Kollman, P. A. Combined Molecular Mechanical and Continuum Solvent Approach (MM-PBSA/GBSA) to Predict Ligand Binding. *Perspect. Drug Discov. Des.* **2000**, *18*, 113–135.
- (200) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods: I. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69–82.
- (201) Crowley, M. F.; Williamson, M. J.; Walker, R. C. CHAMBER: Comprehensive Support for CHARMM Force Fields within the AMBER Software. *Int. J. Quantum Chem.* **2009**, *109*, 3767–3772.
- (202) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (203) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J. Med. Chem.* **2005**, *48*, 4040–4048.
- (204) Pearlman, D. A. Evaluating the Molecular Mechanics Poisson-Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to p38 MAP Kinase. *J. Med. Chem.* **2005**, *48*, 7796–7807.
- (205) Singh, N.; Warshel, A. Absolute Binding Free Energy Calculations: On the Accuracy of Computational Scoring of Protein-Ligand Interactions. *Proteins Struct. Funct. Bioinforma.* **2010**, NA – NA.
- (206) Sono, M.; Roach, M.; Coulter, E.; Dawson, J. (10) Heme-Containing Oxygenases. *Chem. Rev.* **1996**, *96*, 2841–2887.
- (207) Xu, Y.; Colletier, J. P.; Jiang, H.; Silman, I.; Sussman, J. L.; Weik, M. Induced-Fit or Preexisting Equilibrium Dynamics? Lessons from Protein Crystallography and MD Simulations on Acetylcholinesterase and Implications for Structure-Based Drug Design. *Protein Sci. Publ. Protein Soc.* **2008**, *17*, 601–605.

- (208) Capece, L.; Lewis-Ballester, A.; Batabyal, D.; Di Russo, N.; Yeh, S.-R.; Estrin, D. A.; Marti, M. A. The First Step of the Dioxygenation Reaction Carried out by Tryptophan Dioxygenase and Indoleamine 2,3-Dioxygenase as Revealed by Quantum Mechanical/molecular Mechanical Studies. *JBIC J. Biol. Inorg. Chem.* **2010**, *15*, 811–823.
- (209) deVisser, S. P.; Shaik, S.; Sharma, P. K.; Kumar, D.; Thiel, W. Active Species of Horseradish Peroxidase (HRP) and Cytochrome P450: Two Electronic Chameleons. *J. Am. Chem. Soc.* **2003**, *125*, 15779–15788.
- (210) Lonsdale, R.; Houghton, K. T.; Żurek, J.; Bathelt, C. M.; Foloppe, N.; de Groot, M. J.; Harvey, J. N.; Mulholland, A. J. Quantum Mechanics/Molecular Mechanics Modeling of Regioselectivity of Drug Metabolism in Cytochrome P450 2C9. *J. Am. Chem. Soc.* **2013**, *135*, 8001–8015.
- (211) Strickland, N.; Harvey, J. N. Spin-Forbidden Ligand Binding to the Ferrous–Heme Group: Ab Initio and DFT Studies. *J. Phys. Chem. B* **2007**, *111*, 841–852.
- (212) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. 6-31G* Basis Set for Third-Row Atoms. *J. Comput. Chem.* **2001**, *22*, 976–984.
- (213) Pauling, L.; Coryell, C. D. The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. *Proc. Natl. Acad. Sci.* **1936**, *22*, 210–216.
- (214) Weiss, J. J. Nature of the Iron–Oxygen Bond in Oxyhaemoglobin. *Nature* **1964**, *202*, 83–84.
- (215) McClure, D. S. Electronic Structure of Transition-Metal Complex Ions. *Radiat. Res. Suppl.* **1960**, *2*, 218–242.
- (216) Goddard, W. A.; Olafson, B. D. Ozone Model for Bonding of an O₂ to Heme in Oxyhemoglobin. *Proc. Natl. Acad. Sci.* **1975**, *72*, 2335–2339.
- (217) Chen, H.; Ikeda-Saito, M.; Shaik, S. Nature of the Fe–O₂ Bonding in Oxy-Myoglobin: Effect of the Protein. *J. Am. Chem. Soc.* **2008**, *130*, 14778–14790.
- (218) Jensen, K. P.; Roos, B. O.; Ryde, U. O₂-Binding to Heme: Electronic Structure and Spectrum of Oxyheme, Studied by Multiconfigurational Methods. *J. Inorg. Biochem.* **2005**, *99*, 45–54.
- (219) Chen, H.; Lai, W.; Shaik, S. Multireference and Multiconfiguration Ab Initio Methods in Heme-Related Systems: What Have We Learned So Far? *J. Phys. Chem. B* **2011**, *115*, 1727–1742.
- (220) Kepp, K. P. O₂ Binding to Heme Is Strongly Facilitated by NearDegeneracy of Electronic States. *Chemphyschem* **2013**, *14*, 3551–3558.
- (221) Harvey, J. N. Spin-Forbidden Reactions: Computational Insight into Mechanisms and Kinetics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 1–14.
- (222) Oganesyan, V. S.; Sharonov, Y. A. Determination of Zero-Field Splitting and Evidence for the Presence of Charge-Transfer Transitions in the Soret Region of High-Spin Ferric Hemoproteins Obtained from an Analysis of Low-Temperature Magnetic Circular Dichroism. *Biochim. Biophys. Acta* **1998**, *1429*, 163–175.
- (223) Harvey, J. N.; Aschi, M. Modelling Spin-Forbidden Reactions: Recombination of Carbon Monoxide with Iron Tetracarbonyl. *Faraday Discuss.* **2003**, *124*, 129–143.
- (224) Forouhar, F.; Anderson, J. L. R.; Mowat, C. G.; Vorobiev, S. M.; Hussain, A.; Abashidze, M.; Bruckmann, C.; Thackray, S. J.; Seetharaman, J.; Tucker, T.; Xiao, R.; Ma, L.-C.; Zhao, L.; Acton, T. B.; Montelione, G. T.; Chapman, S. K.; Tong, L. Molecular Insights into Substrate Recognition and Catalysis by Tryptophan 2,3-Dioxygenase. *Proc. Natl. Acad. Sci.* **2007**, *104*, 473–478.

- (225) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Inclusion of Dispersion Effects Significantly Improves Accuracy of Calculated Reaction Barriers for Cytochrome P450 Catalyzed Reactions. *J. Phys. Chem. Lett.* **2010**, *1*, 3232–3237.
- (226) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Effects of Dispersion in Density Functional Based Quantum Mechanical/Molecular Mechanical Calculations on Cytochrome P450 Catalyzed Reactions. *J. Chem. Theory Comput.* **2012**, *8*, 4637–4645.
- (227) John, S.; Thangapandian, S.; Sakkiah, S.; Lee, K. W. Identification of Potent Virtual Leads to Design Novel Indoleamine 2,3-Dioxygenase Inhibitors: Pharmacophore Modeling and Molecular Docking Studies. *Eur. J. Med. Chem.* **2010**, *45*, 4004–4012.
- (228) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Effects of Dispersion in Density Functional Based Quantum Mechanical/Molecular Mechanical Calculations on Cytochrome P450 Catalyzed Reactions. *J. Chem. Theory Comput.* **2012**, *8*, 4637–4645.
- (229) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106*, 3210–3235.
- (230) Warshel, A.; Florian, J. Computer Simulations of Enzyme Catalysis: Finding out What Has Been Optimized by Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 5950–5955.
- (231) Steinbrecher, T.; Labahn, A. Towards Accurate Free Energy Calculations in Ligand Protein-Binding Studies. *Curr. Med. Chem.* **2010**, *17*, 767–785.
- (232) Woods, C. J.; Malaisree, M.; Hannongbua, S.; Mulholland, A. J. A Water-Swap Reaction Coordinate for the Calculation of Absolute Protein-Ligand Binding Free Energies. *J. Chem. Phys.* **2011**, *134*, 054114.

