JOURNAL OF PHYSICAL ORGANIC CHEMISTRY
J. Phys. Org. Chem. 2006; **19:** 608–615
Published online 24 April 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/poc.1030

QM and QM/MM studies of selectivity in organic and bioorganic chemistry†

Jeremy N. Harvey,* Varinder K. Aggarwal, Christine M. Bathelt, José-Luis Carreón-Macedo, Timothy Gallagher, Nicole Holzmann, Adrian J. Mulholland and Raphaël Robiette

School of Chemistry, University of Bristol, Cantock's Close, BS8 1TS Bristol, UK

Received 27 September 2005; revised 28 October 2005; accepted 3 November 2005

ABSTRACT: Electronic structure methods are used to explore the origin of selectivity in a number of organic, organometallic and bioorganic processes. Diastereoselectivity in reactions of sulfur and phosphorus ylides is shown to arise during a variety of different elementary steps, and is due to steric and electronic effects. Unusual rearranged products from Heck reaction of *o*-bromo biphenyl derivatives are shown to result from an unusual electrophilic addition step. Predicting selectivity in oxidation of aromatic substrates by cytochrome P450 isoforms is a challenging problem, which can be tackled using hybrid QM/MM methods. Differences in the electronic structure of the Compound I active intermediates of different cytochrome P450 isoforms do not appear to be large enough to explain the different selectivity of these different isoforms. Copyright © 2006 John Wiley & Sons, Ltd.

KEYWORDS: computation; selectivity; QM/MM; Heck reaction; C—H bond activation; sulfur ylides; Wittig reaction; cytochrome P450

INTRODUCTION

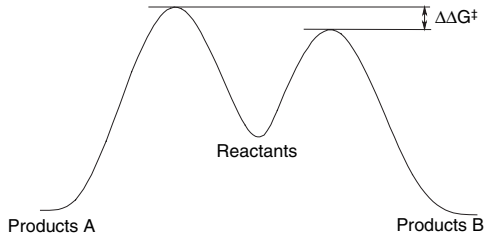
Selectivity is one of the most important objectives of synthetic chemistry and plays an important role in biochemistry also. To be useful, an organic reaction needs to occur at the correct functional group (chemoselectivity), at the right site (regioselectivity), and must yield the desired diastereoisomer (diastereoselectivity) and/or enantiomer (enantioselectivity). To develop a reactive system that meets these criteria, it is very useful to have an accurate understanding of the reaction mechanism, including the knowledge of which steps are responsible for selectivity, and which steric and electronic factors most contribute to selectivity. Of course, in many cases, purely rational design is not practical, but even so, some rough model of the mechanism will often be used as a heuristic guide to the factors that require most care, leading to a guided trial and error procedure. Most biochemical transformations are highly selective, with each enzyme having evolved to transform one particular substrate into one particular product. However, some important types of enzyme do catalyze transformations of a range of compounds and can lead to isomeric products, and understanding the factors influencing the selectivity is then important also. For example, β-lactamases can usually hydrolyze a broad spectrum of β-lactam antibiotics. Reaction with β-lactamase inhibitors follows a similar chemical mechanism yet can crucially lead to inhibition of the enzyme. An understanding of the mechanism can help design pharmaceuticals that will be active against pathogens by selectively leading to inhibition.

How can computational chemistry help in the understanding of selectivity? Increasingly powerful computers and algorithms mean that electronic structure methods are used more and more often to unravel reaction mechanisms. Yet in many cases, a description of the mechanism in some simple model system is not enough to be able to determine the origin of selectivity. For this purpose, it is required to study the pathways leading from the different substrates and to the different isomeric products. Computational chemistry is now able to do this, and thereby provide the required insight, and many reports now appearing in the literature discuss this point. In this account, we will describe a number of recent projects of this type carried out in our group, and discuss the challenges and problems associated with the computational modeling of selectivity. Many other groups have used methods similar to those discussed here to tackle related questions concerning selectivity. Our work is thereby by no means unique, and indeed is in part indebted to the insights provided by others. However, in this overview, we discuss our own recent work and hence do not provide systematic references to work by others.

SELECTIVITY AND ACCURACY

Before discussing our results, it is useful to consider the challenge faced by the computational treatment of

**Correspondence to:* J. N. Harvey, School of Chemistry, University of Bristol, Cantock's Close, BS8 1TS Bristol, UK.
E-mail: jeremy.harvey@bristol.ac.uk
†Selected paper presented at the 10th European Symposium on Organic Reactivity, 20–30 July, Rome, Italy.



Scheme 1

selectivity. Chemical reactions may occur with any sort of proportions of desired and undesired products, but when isomeric products are formed, the corresponding pathways are often very similar and hence close in energy, so that the range of product distributions which are of interest in the context discussed here lies between 1:99 and 99:1. At room temperature, this corresponds to differences in free energy of activation $\Delta\Delta G^{\ddagger}$ (Scheme 1) between the two pathways of ± 3 kcal/mol or less. A ''useful'' level of selectivity – say 90% desired products to 10% by-products – corresponds to an even smaller free energy difference of 1.3 kcal/mol, and if low temperature has been used to enhance selectivity, the same ratio can be obtained with a difference in free energy of below 1 kcal/mol.

This clearly suggests that the direct computational *prediction* of accurate selectivity ratios is almost impossible. Typical *ab initio* and density functional theory (DFT) electronic structure methods aim to reach 'chemical accuracy' on relative energetics, that is, an error of 1 kcal/mol, barely good enough even to predict whether two pathways are equally likely or one is preferred over the other by a factor of 9! On top of this, electronic structure methods lead in a direct way only to gas-phase *energies*, not to condensed-phase *free energies*. Solvent effects are well known to play an important role in selectivity, and gas-phase energy profiles can be meaningless for reactions involving charge separation or recombination. Entropic effects, both for the solutes and the solvent, are well known to have an important effect on rate constants, so that the use of electronic energies rather than free energies is a serious shortcoming.

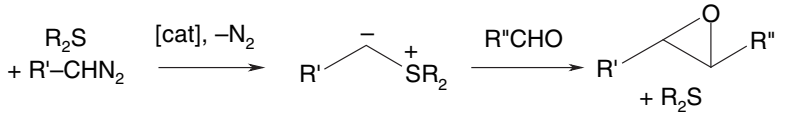
The situation is, however, not as severe as suggested here. First of all, modern methods for including a continuum description of the enthalpic and entropic effects of the solvent have proved to be remarkably successful, and can now be applied routinely.[1] Where one or more solvent molecules is tightly coupled to either reactants, intermediates or transition states, so that the continuum model is inappropriate, it is possible to include these species in the computational model. This has not been necessary in any of the studies discussed below, however. Entropic effects for the solutes can be treated using statistical mechanics, and are moreover expected to be fairly similar for two parallel pathways leading to isomeric products. The main problem to achieving chemical or better accuracy in electronic structure theory is electron correlation, which varies in magnitude depending on the electronic structure of the species considered. Here too, as the key transition states leading to isomeric products are expected to have very similar electronic structures, favorable error cancellation can be expected. For all these reasons, it is realistic to assume that calculated $\Delta\Delta E^{\ddagger}$ or $\Delta\Delta G^{\ddagger}$ values for competing pathways often reach an accuracy of ± 2 kcal/mol or even ± 1 kcal/mol in favorable cases. This level of accuracy is of course not adequate for *predicting* product ratios with high accuracy. However, if calculations lead to energetics that are consistent with observed reactivity and selectivity, then it can be assumed that a good model of the reaction and of the competing pathways has been obtained, and this can be used for insight into developing improved reaction conditions, substituents, etc. The examples below illustrate how this can be achieved using current computational methods for a range of organic reactions. The challenges involved in understanding biochemical selectivity are also illustrated using the example of cytochrome P450 metabolism of xenobiotics.

The calculations described in this work could in principle be carried out using a variety of computational packages. The DFT calculations on molecular systems, including continuum solvent were carried out using the Jaguar program,[2] whereas the QM/MM calculations were carried out using Jaguar (for the QM part), Tinker[3] (for the MM part) and QoMMMa[4] (for the QM/MM coupling).

EPOXIDATION REACTIONS OF SULFONIUM YLIDES

Ylides represent a highly versatile class of organic reactants, which undergo many useful carbon–carbon bond forming processes such as olefination, epoxidation of aldehydes, cyclopropanation, and rearrangement. We have been interested in several of these processes, and discuss here our computations relating to the epoxidation chemistry of sulfonium ylides and to the olefination (Wittig) chemistry of phosphorus ylides.

Sulfonium ylides can be readily generated by formal carbene transfer to a sulfide, and react readily with aldehydes to form epoxides (Scheme 2). Careful optimization of conditions together with the design of



Scheme 2

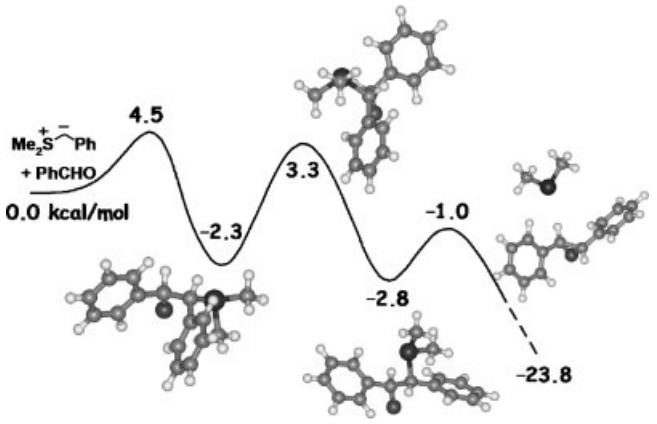


Figure 1. Calculated[6] (B3LYP/6-311 + G**(CH_3CN)//B3LYP/6-31G*(CH_3CN)) energy profile for formation of *trans* stilbene oxide from benzaldehyde and dimethylsulfonium benzylidene ylide

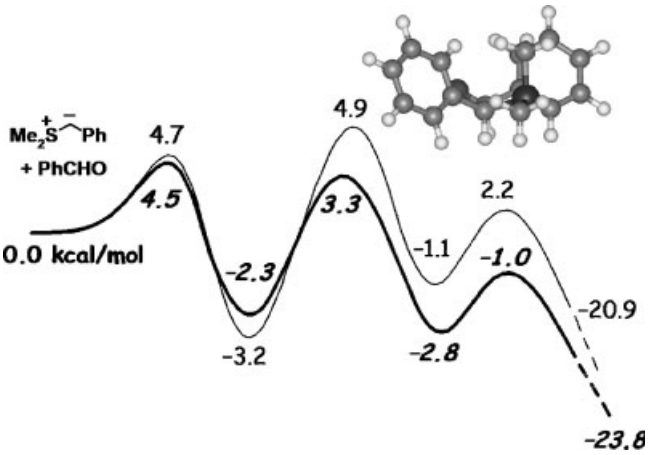


Figure 2. Comparative calculated[6] (B3LYP/6-311 + G** (CH_3CN)//B3LYP/6-31G*(CH_3CN)) energy profiles for formation of *cis* (light line) and *trans* (bold line) stilbene oxide from benzaldehyde and dimethylsulfonium benzylidene ylide

suitable chiral sulfides has enabled this reaction to be developed into a very general tool for the enantioselective synthesis of a broad range of epoxides.[5] To assist in this development, it was necessary to obtain better insight into the mechanism of the process and the identity of the steps leading to diastereoselectivity between *cis* and *trans* epoxides and to enantioselectivity.

We have first of all studied the important case of semi-stabilized ylides, in which R′ is a phenyl or other aryl group.[6] Our computations addressed the simplest case of the dimethylsulfide-derived ylides, in an attempt to understand the observed preference for *trans* epoxides. The computed energy profile for the pathway leading to the *trans* epoxide is shown in Fig. 1.[7]

As can be seen, the reaction involves three steps. First, addition of the ylide occurs to form the new carbon–carbon bond. The intermediate thus created is not a cyclic oxathietane (analogous to the oxaphosphetane intermediates of Wittig reactions), but a betaine. In principle, this type of nucleophilic addition could occur to give either 'transoid' or 'cisoid' betaines, that is, the oxy and dimethylsulfonium groups could lie either roughly eclipsed or *anti* to one another. In practice, we found that the transition state leading to the cisoid betaine shown was always somewhat lower than that leading to the rotamer, presumably due to electrostatic attraction between the positive and negative charges. The betaine is very slightly more stable in energy terms than the reactants, but due to the fact that two molecules have combined to give one, there will be a loss of entropy so that the betaine will lie higher in free energy than the reactants and hence should not build up and thereby be observable during reaction. The next step is torsional rotation around the new C—C bond. This type of rotation is usually very easy, and the computed barrier in the present case is low also, at only 5.6 kcal/mol relative to the initial betaine. However, the other barriers along the reaction pathway are low also, so this step cannot be neglected. Finally, the transoid betaine produced by this conformational change is able to undergo intramolecular substitution leading to the products, over a very low barrier.

Note that it is only possible to produce a sensible reaction profile for this process by optimizing geometries and calculating energies in the presence of a continuum solvent model. This is because the polarity and hence solvation energy of the different species is highly variable: the reactants are polar but neutral, the cisoid betaine is zwitterionic but with two relatively closely positioned charges, the transoid betaine is extremely sensitive to solvation because the two charges are now separated, and the products are almost non-polar. In the gas phase, attempted optimization of the transoid betaine leads directly to products.[6]

We have also calculated the energy profile leading to the minor *cis* stilbene oxide (Fig. 2). The overall sequence of steps is the same in both cases, but the relative energies of the transition states are different. Whereas addition is the highest barrier for formation of the *trans* product, torsional rotation is the highest point along the pathway leading to *cis* products. Taking the relative height of the calculated barriers at face value, addition is expected to be non-reversible when it produces the *anti* betaine precursor of the *trans* products, but reversible when it leads to the diastereoisomeric *syn* betaine. This is consistent with the observed high *trans* selectivity of the overall reaction. It is also consistent with crossover experiments.[8] The betaines can be produced independently by deprotonation of the β-hydroxy sulfonium salts. When this is done in presence of a reactive aldehyde such as *p*-nitrobenzaldehyde, the *anti* betaine leads to *trans* stilbene oxide primarily, whereas the *syn* betaine leads to a mixture of *cis* and *trans* stilbene oxide and the nitrated derivative. This shows that the betaine has undergone some reversion to the ylide, which has then added back to the aldehydes present in solution to yield *syn* and *anti* betaines.

It can be seen that the energy differences between the calculated barriers are very small – the barrier to reversion

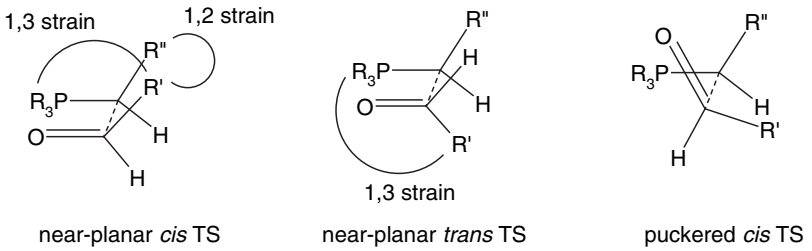


Figure 3. Transition state models for initial cycloaddition step in Wittig reactions

for the *syn* cisoid betaine (4.7 kcal/mol relative to reactants, 7.9 kcal/mol relative to the betaine) is barely lower than that for torsional rotation (4.9 kcal/mol relative to reactants). It is, therefore, not possible to *deduce* from these calculations that reactivity and selectivity will be as described above – the calculations are simply not accurate enough. However, the calculated surfaces do suggest a pattern of reactivity that is consistent with experimental observations. They also identify the key step that leads to diastereoselectivity: torsional rotation. This was unexpected, as this kind of step is usually much easier than any other. This shows how computation can provide additional insight into the origin of stereoselectivity. For example, the suggested mechanism clearly shows that with chiral sulfide precursors to the ylides, the enantioselectivity-defining step is the initial addition to give *anti* cisoid betaine, but that this pattern could be quite easily disrupted if substituents on the ylide or the aldehyde made torsional rotation rate-limiting also for the *anti* pathway leading to *trans* epoxides. This analysis has been used to understand the observed trends in enantio- and diastereoselectivities with various chiral sulfides.[9]

We have also studied the reaction of more highly stabilized sulfur ylides R_2S^+—CH^-—$CONR'_2$. Interestingly, our calculations[10] show that in this case, the electron-withdrawing group slows down the intramolecular ring-closing step, such that it becomes rate limiting in both the *syn* and *anti* pathways. The origin of enantioselectivity and diastereoselectivity is thereby completely different from that in the reaction of the semi-stabilized ylides R_2S^+—CH^-—Ph.

WITTIG REACTIONS OF STABILIZED PHOSPHONIUM YLIDES

Recently we have made some surprising observations[11] concerning the *E/Z* selectivity of Wittig reactions of novel ylides derived from phosphites, $(RO)_3P^+$—CH^-—R′, where R′ is an aryl (semi-stabilized ylide) or an ester (stabilized ylide) group. With more common triphenylphosphonium ylides, the semi-stabilized ylides are known to yield low *E/Z* selectivity, whereas the stabilized ylides give high *E* selectivity. The opposite trend is found with the phosphite-derived ylides. The behavior of the triphenylphosphonium ylides can be rationalized in terms of the widely accepted models suggested by Vedejs et al., as shown in Fig. 3.[12]

It is known[12] that in most cases, in the absence of metal ions, the reaction involves an initial rate- and selectivity-determining addition step leading to *cis* or *trans* oxaphosphetanes, followed by selective loss of phosphine oxide to give the corresponding *Z* or *E* alkenes. When the initial addition occurs through a near-planar transition state, *trans* oxaphosphetane and hence *E* alkene is favored, because the *cis* addition TS is destabilized by 1,2 interactions between the R′ group of the alkene and the R″ group of the ylide. Both TSs are destabilized to a similar extent by 1,3 interactions between the aldehyde R′ group and the bulky phosphonium residue. This is the situation for stabilized ylides, which react through late, near-planar TSs. For non-stabilized ylides, however, the TS is quite early and hence fairly flexible, and the *cis* TS can undergo a puckering deformation which relieves both 1,2 and 1,3 strain. A similar deformation of the *trans* TS relieves 1,3 strain but at the cost of increased 1,2 interactions, and *Z* alkenes are favored. Semi-stabilized ylides fall somewhere in between, so that the low selectivity is not unexpected.

Phosphite derived ylides are more strongly stabilized than analogous triphenyl phosphonium compounds, due to delocalization of the ylidic negative charge into the P—O σ^* orbitals. This explains the observed selectivity obtained with the 'semi-stabilized' aryl derivative: despite the absence of an electron-withdrawing group on the ylidic carbon, such ylides are in fact stabilized, and so the TS model of Fig. 3 predicts late addition TSs and high E selectivity, as is observed experimentally.[11] On the other hand, the low *E* selectivity obtained[11] with the even more stabilized ylides $(RO)_3P^+$—CH^-—CO_2R', (E/Z typically only 70:30 with PhCHO) is surprising and prompted us to examine the pathways computationally.

Our calculations[13] on the stabilized ylides of both types (at the B3LYP/6-31G*(THF) level) confirm that the initial addition is rate limiting, and is followed by facile elimination of phosphine oxide. We also find that the *trans* addition TS is significantly (ca. 3 kcal/mol) lower in energy than the *cis* one in the case of the triphenylphosphonium ylide, consistent with the experimental observation of high *E* selectivity in this case. The difference in energy between the two TSs derived from the phosphite-based ylides is much smaller, again consistent with the low *E* selectivity observed with those species.

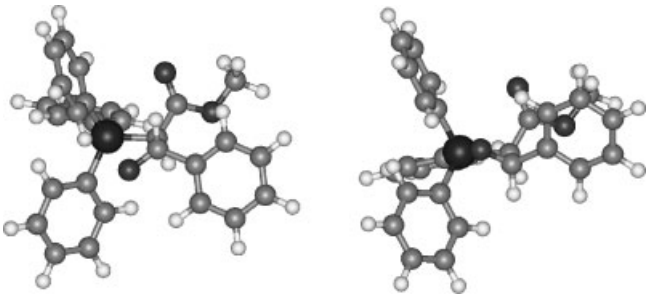


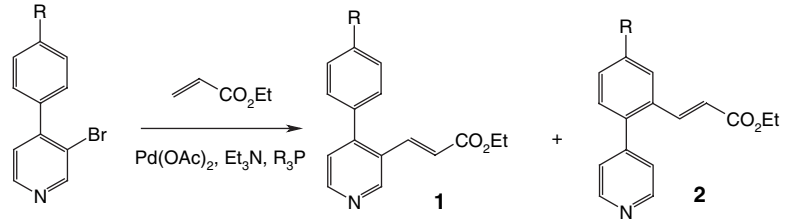
Figure 4. Structure of optimized *trans* (left) and *cis* (right) TSs (B3LYP/6-31G*) for addition of Ph_3PCHCO_2Me to PhCHO

The explanation for these energy trends is interesting, and is best grasped by considering the structures of the *trans* and *cis* TSs for the triphenylphosphonium reaction. As can be seen in Fig. 4, the *trans* TS is significantly puckered (PCCO dihedral angle of −40°), whereas the *cis* TS is closer to planarity (PCCO dihedral angle of 26°). These geometries are thereby doubly surprising: the model of Fig. 3 suggests that both TSs should be close to planar, with the *cis* TS having most to gain in terms of relief of steric strain from puckering. Clearly, some previously unrecognized effect must be in operation here, and careful analysis shows that this is electrostatic (dipole–dipole) interaction between the dipoles of the two reactants at the TSs. The aldehyde reactant has a strong dipole (3.37 *D*) along the C═O bond, and the ylide has an even larger dipole (5.32 *D*) oriented along the C_{ylidic}—C_{ester} bond, due to delocalization of the ylidic negative charge into the ester group. The puckering of the *trans* TS enables these two dipoles to adopt a near-ideal antiparallel arrangement, whereas the near-planar structure of the *cis* TS results from an attempt to minimize repulsion between the dipoles which would be near-parallel in case of puckering. Selectivity thereby results from a combination of unfavorable steric and electrostatic interactions in the *cis* TS. The lower selectivity with the phosphite-derived ylides is due to the fact that the dipole in that case is much smaller (2.12 *D*), because of competing delocalization of the negative charge into the ester and trialkoxyphosphonium groups.[13]

CROSSOVER PRODUCTS IN THE HECK REACTION

In recent experimental work in Bristol,[14] unexpected crossover products **2** were observed in the Heck reactions of 4-aryl 3-bromopyridines (Scheme 3). Formation of these products formally involves activation of a C—H bond in the ortho position of the aryl substituent.

The mechanism suggested in the experimental work (Fig. 5) involves initial insertion of Pd(0) into the C—Br bond to yield **3**, which can then either undergo normal Heck chemistry to give **1**, or undergo insertion of Pd into the C—H bond combined with loss of HX to form a palladacycle **4** (Fig. 5). Species such as **4** are known and are able to form Heck products.[15] We are interested in understanding how **4** can be formed, and how to account for the observed selectivity between normal products **1** and crossover products **2** depending on substituents.



Scheme 3

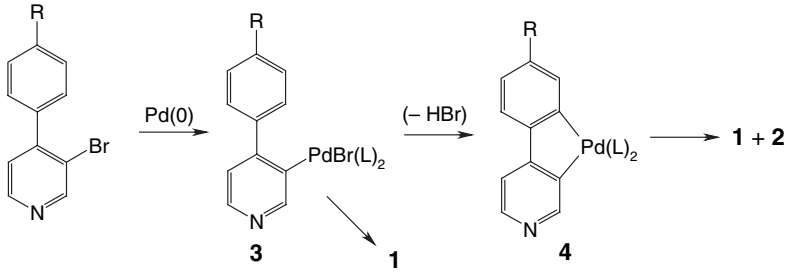


Figure 5. Suggested mechanism leading to crossover products through palladacycle **4**

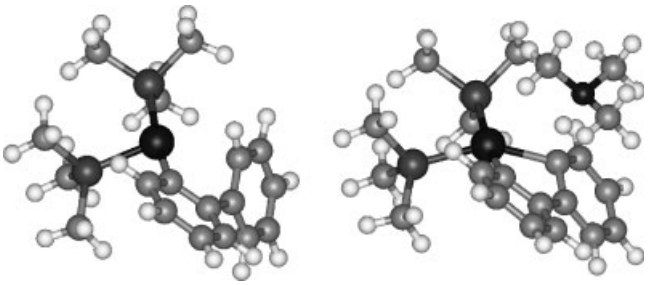
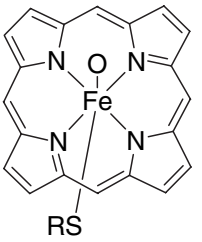


Figure 6. Cationic intermediate (left) and deprotonation TS (right) leading to palladacycle formation



3

Scheme 4

In our calculations,[16] we use trimethylphosphine as a model ligand, and the B3PW91 level of theory together with a continuum model of the acetonitrile solvent. The bulk of our calculations have been carried out on an all-carbon biphenyl system instead of the pyridyl system shown in Scheme 3 and Fig. 5. We first of all considered mechanisms involving Pd(IV) intermediates in which the Pd center in **3** has undergone insertion into the *ortho* C—H bond either directly (yielding various geometric isomers of an octahedral Pd(IV) species) or after initial loss of phosphine or bromide. However, none of these species are likely to be formed as intermediates due to their high energy (more than 40 kcal/mol above **3**). In some cases, they could not even be located as minima, and revert instead to Pd(II) species with an intact C—H bond upon geometry optimization.

Instead, we find a low-energy pathway involving rate-limiting deprotonation of the cationic intermediate formed after dissociation of bromide from **3**. As can be seen in Fig. 6, this species involves a fairly close contact between the cationic Pd and the ortho carbon atom ($r_{Pd-C} = 2.56$ Å). The ortho hydrogen atom can be removed by the base present under Heck conditions (modeled here as trimethylamine), and the optimized TS for this step is also shown in Fig. 6. As the proton is removed, the carbon atom can directly bond to Pd. Indeed, r_{Pd-C} decreases to 2.19 Å at the TS, already close to its value of 2.05 Å in the palladacycle. Dissociation of bromide is endothermic by 11 kcal/mol in the presence of a continuum model of acetonitrile, and the TS lies a further 13.5 kcal/mol higher. Given that the bromide dissociation is entropically favored, this produces a fairly modest overall activation free energy, consistent with the observation that this process competes with 'normal' Heck chemistry. In the real system, the phosphine ligand is triphenyl phosphine or tris(*o*-tolyl)phosphine, so the real energetics may be somewhat different to those for the present model trimethylphosphine system.

The mechanism outlined above is formally equivalent to an electrophilic substitution on the aryl ring. However, the palladium atom does not add to the ring before proton loss, and this changes the selectivity. As shown in Scheme 3, crossover occurs to a greater extent when an *electron-withdrawing* substituent is present on the second ring—the opposite trend to that expected for normal electrophilic substitution. However, given that electron-withdrawing substituents can be expected to increase the acidity of the ortho proton, our mechanism is consistent with the observed selectivity. Calculations on systems where the aryl group is varied also agree with the observed trend. The energy barriers for the deprotonation step are of respectively 13.7, 13.5, 12.2, and 10.6 kcal/mol when the aryl group is *p*-methoxyphenyl, phenyl, 4-pyridyl and *p*-nitrophenyl. This decrease in the barrier as the electron-withdrawing character increases is consistent with the observed increased ratio of crossover products formed.

REGIOSELECTIVITY IN BIOORGANIC OXIDATION

The cytochrome P450 enzymes catalyze the oxidation of strong aliphatic and aromatic C—H bonds in a variety of organic compounds, for example, drugs in the human liver.[17] We are interested in understanding the chemo-selectivity of oxidation of different substrates by different isoforms, as well as the regio- and stereo-selectivity of oxidation of C—H bonds for a given substrate and isoform.

One aspect of this selectivity is due to intrinsic electronic factors, associated with the mechanism of reaction of the key intermediate, Compound I (**3**, Scheme 4). We have carried out gas-phase calculations of the barrier to oxidation of C—H bonds in aromatic compounds, using a model of the active species, and a variety of monosubstituted aromatics.[18] The rate-limiting step involves addition of Compound I to the aromatic ring to form a covalently bound adduct that subsequently undergoes facile rearrangement to form products. We found that both electron-withdrawing and electron-donating substituents on the *para* position of the aromatic ring activate the ring for C—O bond formation, due to the mixed electrophilic-radical nature of the addition step. The calculated energy barriers could be reproduced using a structure-activity relation based on a combination of radical and cationic Hammett substituent parameters.

These calculations address the *intrinsic* reactivity of the different positions of an aromatic ring, for a simple model of the common active species of all isoforms of cytochrome P450. Such calculations cannot yield full insight into the selectivity of these enzymes, as the detailed protein environment clearly plays a role also. For example, different isoforms can oxidize a substrate at different positions, as in the case of the anti-inflammatory diclofenac, which undergoes hydroxylation on one of its two aromatic rings in the 2C9 isoform, and on the other in the 3A4 isoform.[19] To understand such effects, it is necessary to use a more complete model, treating a large sphere of atoms around the active site. This is done most conveniently using hybrid quantum mechanical/molecular mechanical (QM/MM) methods[20] in which a subset of atoms are treated using an explicit quantum mechanical method to describe their electronic structure, whereas the bulk of the protein and the surrounding water molecules are treated using much simpler molecular mechanics.

Potential energy surfaces for biomolecular systems are extremely complicated, as reactant and product complexes, as well as transition states, will typically exist in many different conformers associated with librational motion of solvent water and with rotation around single bonds of amino-acid side chains. The activation free energy for a reaction arises from thermal averaging over the ensemble of conformers corresponding to reactants and that corresponding to the transition state. It is not straightforward to carry out the necessary averaging using expensive QM/MM methods, yet activation energies taken as differences in energy between a particular conformer of the reactant complex and another of the TS may not be meaningful. In this context, calculating selectivity effects may appear to be impossible. However, consideration of reaction pathways corresponding to multiple conformers of reactant complexes is capable of providing accurate activation barriers, as shown in our recent work on the enzyme chorismate mutase, where the experimental activation enthalpy is 12.7 kcal/mol, and the average computed activation barrier for a number of different conformers is 12.0 kcal/mol.[21]

Also, in one favorable case, some of us have shown that computed activation energies for hydroxylation of a series of phenolic substrates by the enzyme phenol hydroxylase correlate with observed reactivity.[22] The correlation (which has a high correlation coefficient of 0.98) is shown in Fig. 7. This remarkable result, given the complexity of the system, is no doubt due to very favorable error cancellation, which enables the experimental trend to be correctly reproduced. In turn, this is due to the fact that a same conformer of the reactant complex was used for all the substrate molecules considered, to the high steric similarity of the latter (all halogenated phenols), and the consistent binding mode for all substrates, associated with the anionic nature of the substrates (which bind in the phenolate form). Accordingly, it may not always be so easy to predict

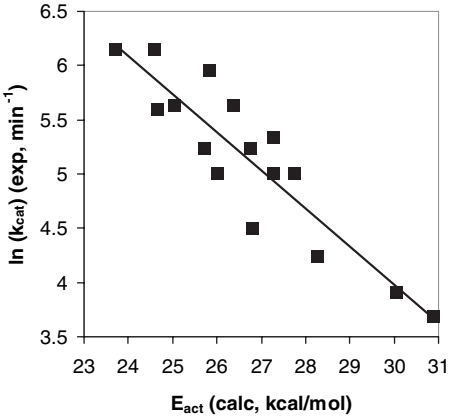


Figure 7. Correlation between measured rate constants and calculated activation barriers for hydroxylation of a range of substituted phenols by the enzyme phenol hydroxylase. (See Ref. 22)

selectivity. However, the success in this special case shows that it is not in principle impossible to explore selectivity effects even in complex biological systems.

In the case of cytochrome P450, we are currently exploring reaction pathways for oxidation of realistic drug molecules such as diclofenac, in an effort to understand selectivity at the atomistic level. In this case, however, the mode of binding of the substrate in the active site is less systematic from one substrate to another, or from one binding mode to another, given the lack of an 'anchoring' polar group which forces the substrate to adopt a particular orientation.

One alternative effect, which may play an important role in the selectivity of cytochrome P450 isoforms, is changes in the electronic structure and hence in the reactivity of the Compound I intermediate due to differences in the protein environment. Compound I is a very reactive species, containing iron in a very high formal oxidation level of +V. In fact, this high oxidation level cannot be sustained, and the iron is in fact known from studies of related Compound I species to be in the +IV oxidation state, with the porphyrin ring of the heme group partly oxidized, with one electron removed from a high-lying a_{2u}-like orbital. Computational studies on Compound I show that this unpaired electron can actually delocalize over the proximal sulfur ligand of the iron atom as well, and that the predicted electronic structure varies quite strongly depending on the environment.[23] In this sense, Compound I is a 'chameleon'. Using QM/MM calculations, we have found that quite subtle differences in the protein environment cause related even larger changes between the electronic structure of the Compound I species of cytochrome c peroxidase (CcP) and ascorbate peroxidase (APX).[24] In APX, the unpaired electron resulting from formal reduction of the iron center from +V to +IV resides on the porphyrin ring, as in the cytochrome P450 case. In CcP however, the unpaired electron is located on a tryptophan residue situated close to the heme group. These differences have been

characterized experimentally, and our calculations reproduce them. This suggests that the QM/MM method is able to describe the subtle polarization effects of the protein environment on Compound I intermediates.

We therefore set out to explore the electronic structure of the Compound I intermediate in a number of different isoforms of mammalian cytochrome P450 enzymes, focusing on members of the three important drug-processing families 3A, 2B, and 3C. Our QM/MM calculations[25] at the B3LYP level of theory find that Compound I does behave like a chameleon: small changes in the protein structure associated with differences in the set-up of the QM/MM calculations, MD equilibration, or the use of a different X-ray crystallographical starting structure leads to a different balance of unpaired electron density on the porphyrin ring and the proximal sulfur ligand. However, these changes are not systematic: the difference in electronic structure from one calculation on a given isoform to another is as large as those between calculations on different isoforms. Hence, it does not appear that the origin of selectivity in different isoforms of cytochrome P450 is due to differences in electronic structure of the key intermediate.

CONCLUSIONS

The examples given above show that computational methods can yield insight into the origin of selectivity in chemical reactions. To do so, it is necessary to predict relative activation free energies for competing pathways to within an accuracy of roughly 1–2 kcal/mol. The selected examples suggest that this is possible, presumably due to favorable error cancellation between the two very similar transition states involved in competitive steps. The accuracy is not high enough to be able to analyze smaller selectivity effects (e.g., the difference in activation free energies for a chiral auxiliary yielding an enantiomeric excess of 92% and another giving 96% can certainly *not* be distinguished based on calculations of the type discussed here). Also, great care is needed in interpreting the results and in making sure that the model used for the reaction is meaningful. The main benefit from such calculations is that they can help revise and improve the heuristic models used by experimental chemists to analyze existing results and design new systems. The work described here provides several examples of this fruitful collaboration of computation and experiment.

Acknowledgements

The authors thank EPSRC (JNH, Advanced Research Fellowship), Merck and Pfizer (VKA, unrestricted grants), the Mexican CONACYT (JLCM), the Université catholique de Louvain (RR, F.S.R. grant), Vernalis plc (AJM), the IBM High Performance Computing Life Sciences Outreach Programme (AJM), and BBSRC (AJM) for their support.

REFERENCES

1. Typical continuum solvent models, as used e.g. in many of our calculations discussed below, provide an estimate of the *free energy* of solvation of the quantum-mechanical solute, thereby including *en gros* solvent entropy effects.
2. Jaguar 5.0, Schrodinger, Inc.: Portland, OR, 1991–2003.
3. Ponder JW. *TINKER: Software Tools for Molecular Design, v4.0*: Saint Louis, MO, 2003.
4. Harvey JN. *Faraday Discuss*. 2004; **127**: 165–177. DOI: 10.1039/b314768a.
5. Aggarwal VK, Winn CL. *Acc. Chem. Res.* 2004; **37**: 611–620. DOI: 10.1021/ar030045f
6. Aggarwal VK, Harvey JN, Richardson J. *J. Am. Chem. Soc*. 2002; **124**: 5747–5756. DOI: 10.1021/ja025633n
7. This 'energy' profile (as well as that shown in Fig. 2) includes the potential energy terms for the reacting system, as well as the free energy of solvation. It is therefore a hybrid of energy and free energy terms which is in many ways similar to the frequently used gas-phase potential energy profiles. Note also that here as elsewhere, we cite more significant figures than are strictly warranted based on the expected errors involved in the method. This is because the numbers as obtained from the calculations are quite *precise* (that is, not *accurate* compared to experiment, but *reproducible* with the number of significant figures quoted) and hence are useful for other authors seeking to reproduce some of our work.
8. Aggarwal VK, Calamai S, Ford JG. *J. Chem. Soc., Perkin Trans. 1*, 1997; 593–599. DOI: 10.1039/a606925h
9. Aggarwal VK, Richardson J. *Chem. Comm.* 2003; 2644–2651. DOI: 10.1039/b304625g
10. Aggarwal VK, Fuentes D, Harvey JN, Hynd G, Ohara D, Picoul W, Robiette R, Smith C, Vasse JL, Winn CL. *J. Am. Chem. Soc.* 2006; **128**: 2105–2114. DOI: 10.1021/ja0568345
11. Aggarwal VK, Fulton JR, Sheldon CG, de Vicente J. *J. Am. Chem. Soc*. 2003; **125**: 6034–6035. DOI: 10.1021/ja029573x
12. For reviews on mechanistic aspects of the Wittig reaction, see (a) Vedejs E, Peterson MJ. *Top. Stereochem*. 1994; **21**: 1–157; (b) Vedejs E, Peterson MJ. In *Advances in Carbanion Chemistry*, vol. 2, Snieckus V (ed). JAI Press: Greenwich, CN, 1996.
13. Robiette R, Richardson J, Aggarwal VK, Harvey JN. *J. Am. Chem. Soc*. 2005; **127**: 13468–13469. DOI 10.1021/ja0539589. Robiette R, Richardson J, Aggarwal VK, Harvey JN. *J. Am. Chem. Soc.* 2006; **128**: 2394–2409. DOI: 10.1021/ja056650q
14. Karig G, Moon MT, Thasana N, Gallagher T. *Org. Lett*. 2002; **4**: 3115–3118. DOI: 10.1021/ol026426v
15. Satoh T, Jones WD. *Organometallics* 2001; **20**: 2916–2919.
16. Holzmann N, Carreon-Macedo JL, Harvey JN, Gallagher T. To be submitted.
17. For a review, see Meunier B, de Visser SP, Shaik S. *Chem. Rev.* 2004; **104**: 3947–3980. DOI: 10.1021/cr020443g
18. (a) Bathelt C, Ridder L, Mulholland AJ, Harvey JN. *J. Am. Chem. Soc*. 2003; **125**: 15004–15005. DOI: 10.1021/ja035590q; (b) Bathelt C, Ridder L, Mulholland AJ, Harvey JN. *Org. Biomol. Chem*. 2004; **2**: 2998–3005. DOI: 10.1039/b410729b
19. Mancy A, Antignac M, Minoletti C, Dijols S, Mouries V, Duong NTH, Battioni P, Dansette PM, Mansuy D. *Biochemistry* 1999; **38**: 14264–14270. DOI: 10.1021/bi991195u
20. Warshel A, Levitt M. *J. Mol. Biol*. 1976; **103**: 227–249. DOI: 10.1016/0022-2836(76)90311-9
21. Claeyssens F, Ranaghan KE, Manby FR, Harvey JN, Mulholland AJ. *Chem. Commun*. 2005; 5068–5070. DOI: 10.1039/b508181e
22. Ridder L, Mulholland AJ, Rietjens IMCM, Vervoort J. *J. Am. Chem. Soc*. 2000; **122**: 8728–8738. DOI: 10.1021/ja0007814
23. See e.g. (a) Schöneboom JC, Lin H, Reuter N, Thiel W, Cohen S, Ogliaro F, Shaik S. *J. Am. Chem. Soc*. 2002; **124**: 8142–8151. DOI: 10.1021/ja026279w; (b) de Visser SP, Shaik S, Sharma PK, Kumar D, Thiel W. *J. Am. Chem. Soc*. 2003; **125**: 15779–15788. DOI: 10.1021/ja0380906
24. Bathelt CM, Mulholland AJ, Harvey JN. *Dalton Trans*. 2005; 3470–3476. DOI 10.1039/b505407a
25. Bathelt CM, Zurek J, Mulholland AJ, Harvey JN. *J. Am. Chem. Soc*. 2005; **127**: 12900–12908. DOI 10.1021/ja0520924