



## Computing Organic Stereoselectivity – from Concepts to Quantitative Calculations and Predictions

Journal:	<i>Chemical Society Reviews</i>
Manuscript ID	CS-REV-07-2016-000573.R1
Article Type:	Tutorial Review
Date Submitted by the Author:	03-Sep-2016
Complete List of Authors:	Peng, Qian; Oxford, Chemistry Research Laboratory Duarte, Fernanda; Oxford, Chemistry Research Laboratory Paton, Robert; Oxford, Chemistry Research Laboratory



Chem Soc Rev

## TUTORIAL REVIEW

## Computing Organic Stereoselectivity – from Concepts to Quantitative Calculations and Predictions

Received 00th January 20xx,  
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Qian Peng<sup>a,b,†</sup>, Fernanda Duarte<sup>a,b,†</sup> and Robert S. Paton<sup>\*a,b</sup>

Advances in theory and processing power have established computation as a valuable interpretative and predictive tool in the discovery of new asymmetric catalysts. This tutorial review outlines the theory and practice of modeling stereoselective reactions. Recent examples illustrate how an understanding of the fundamental principles and the application of state-of-the-art computational methods may be used to gain mechanistic insight into organic and organometallic reactions. We highlight the emerging potential of this computational tool-box in providing meaningful predictions for the rational design of asymmetric catalysts. We present an accessible account of the field to encourage future synergy between computation and experiment.

### Key learning points

- (1) Simple and useful models to rationalise the origins of enantioselectivity can be obtained through mechanistic understanding and the use of affordable computational methods.
- (2) A deeper understanding of the electronic effects underlying such selectivity can be reached by topology and orbital-based analyses.
- (3) Conformational space sampling and inclusion of explicit solvent effects remain challenging issues.
- (4) Quantitative validation against experiments is necessary to ensure that computational results are “right for the right reason”.

### 1. Introduction

The search for stereoselective chemical reactions is driven both by fundamental academic curiosity as well as society's need for new pharmaceuticals, agricultural chemicals, and materials. The development of asymmetric catalysts, in which each molecule of a chiral catalyst can yield many molecules of chiral product, continues to challenge our understanding of the factors controlling rate and selectivity. Traditionally, however, mechanistic and computational insight has followed rather than led synthesis. This is beginning to change: computational insights, which contribute to the theoretical understanding of asymmetric catalysis and enable predictions to be made, will play an important role in the development of new catalysts.

Asymmetric transition metal catalysis and computational chemistry both emerged coincidentally as prominent areas of research towards the end of the 20<sup>th</sup> century. Chemistry Nobel Prizes were awarded to pioneers of theory and computation Kohn and Pople in 1998, and then in 2001 to Knowles, Noyori and

Sharpless for the discovery of asymmetric catalytic methods in synthesis. Both fields continue to grow apace, for example through the discovery of new metal-free and organocatalytic asymmetric reactions. During these developments it is fair to say that computational chemistry has, in the main, been used as an interpretative rather than a predictive tool. However, with the advent of new theoretical methods and exponential increase in computational power, theory and computation have become more active, and their predictive power exploited even more<sup>1,2</sup>. This can be seen in the increasing number of research works combining both experiment and computation (Fig. 1), which have contributed to a better understanding of asymmetric catalytic processes and most recently to the discovery of new catalysts.

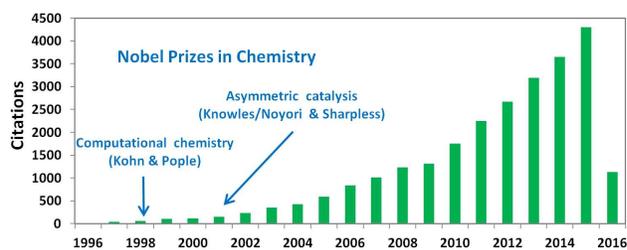


Fig. 1 Web of Science™ citation counts on 26 April 2016 associated with keywords of TOP=(Comput\* AND asymmetric AND Catalys\*).

In this tutorial review, we outline contemporary computational techniques that have become a popular tool for rationalising and

<sup>a</sup> Chemistry Research Laboratory, Department of Chemistry, University of Oxford, 12 Mansfield Road, Oxford OX1 3TA, U.K.

<sup>b</sup> Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ, U.K.

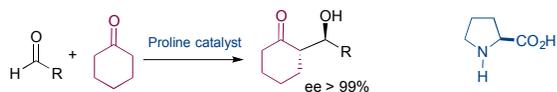
\* Corresponding author: robert.paton@chem.ox.ac.uk

† These authors contributed equally to this work.

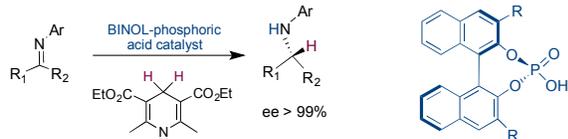
predicting the stereoselectivity of chiral reagents and catalysts. First, we briefly introduce the basic principles underlying stereoselectivity

Building on these principles we then discuss the different ways in which experimentalists and computational chemists quantify selectivity, and show how they can be obtained from widely available programs. Prominent examples have been chosen to demonstrate how computational chemistry has been used to gain a mechanistic understanding of these processes, leading to working models used by synthetic and computational chemists (Scheme 1).

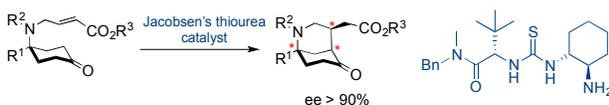
#### a) Proline-catalysed aldol reaction



#### b) Binol-derived phosphoric acids as Brønsted acid catalysts



#### c) Thiourea-catalysed Michael addition reaction



**Scheme 1:** Examples of catalytic enantioselective processes discussed in this review: a) proline, b) BINOL phosphoric acids and c) thioureas.



Qian Peng



Fernanda Duarte



Robert S. Paton

Qian Peng received his PhD in organic chemistry from the Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences. During his postdoctoral working at the University of Notre Dame, the focus of his work shifted to computational spectroscopy. In 2013, Qian moved to Texas A&M University for the research of computational inorganic chemistry and in 2014 he was awarded as Marie Curie Fellow working at the University of Oxford. His research interests focus on applying computational techniques to solve chemical problems in spectra prediction, atom economic reactions catalytic processes and chiral control.

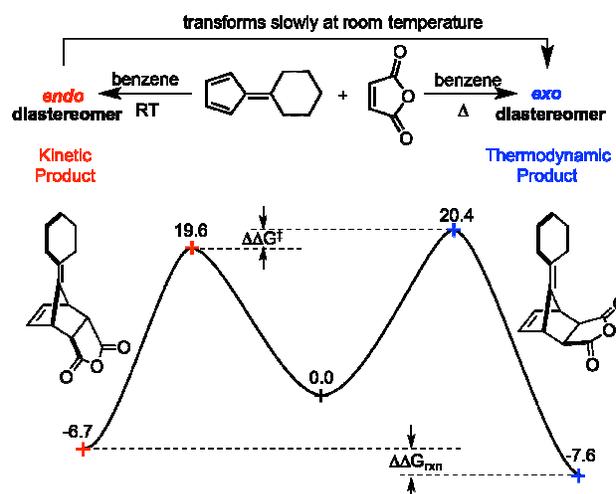
Fernanda Duarte studied chemistry at Pontificia Universidad Católica de Chile, obtaining her PhD in 2012. Shortly after this, she became a Postdoctoral researcher at Uppsala University, applying physical organic chemistry and computational techniques to study phosphoryl and sulfur transfer reactions in solution and in enzymes. She is currently a Royal Society Newton Research Fellow at University of Oxford. Her research applies computational chemistry to characterize and redesign organocatalytic processes. She has received numerous recognitions, including the L'Oréal Unesco "Women for Science" award and a fellowship from the Bert and Kuggie Vallee Foundation to attend the Lindau Nobel Laureate Meeting.

Robert Paton is from Stockport (born 1981) in the North of England and completed undergraduate and graduate studies at the University of Cambridge. From 2008 to 2009, he worked at ICIQ, Spain and at UCLA, USA as a postdoctoral associate, Fulbright Scholar and Fellow of the Royal Commission of 1851. In 2010 Rob was appointed to a University Lectureship in Organic Chemistry at the University of Oxford, where he is currently Associate Professor and Tutorial Fellow. His research interests focus on the elucidation of organic mechanisms and structures with computation.

This tutorial concludes by highlighting recent advances in the field as well as outstanding challenges. We hope this tutorial review will encourage and assist non-specialists to apply computational modelling in synergy with experiment.

## 2. Kinetic vs Thermodynamically Controlled Stereoselectivity

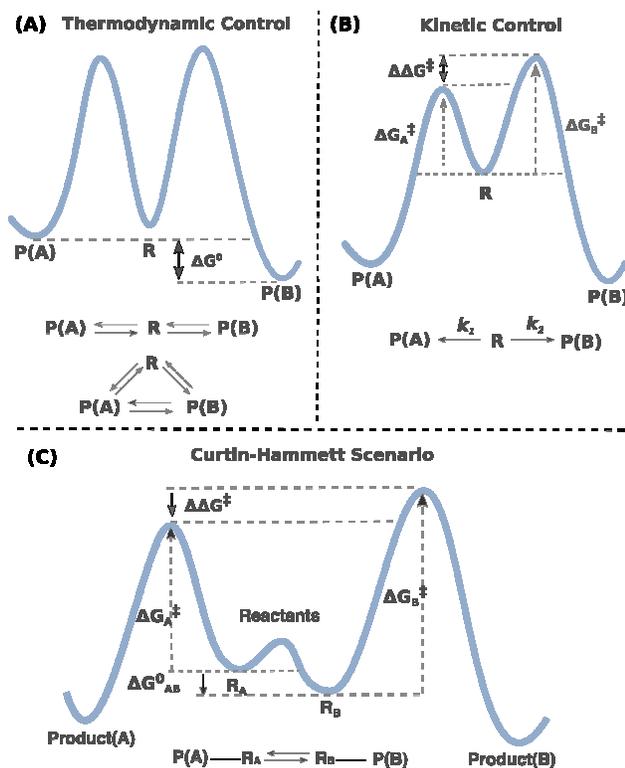
In 1944 Woodward and Baer reported the Diels–Alder reaction between 6,6-pentamethylenefulvene and maleic anhydride (Scheme 2)<sup>3</sup>. The observed reaction led to a mixture of *endo*- and *exo*- diastereomers, initially described as  $\alpha$ - and  $\beta$ -adducts, either of which could be favoured by changing the experimental conditions. For example, while the *endo*-adduct was formed in greater amount at low temperature, the *exo*-adduct became predominant at higher reaction temperatures or after allowing the mother liquor to stand for several weeks. Based on the concepts of transition state (TS), activation barrier, and potential energy surface (PES), introduced earlier by Eyring, Evans and Polanyi, a qualitative potential energy diagram was used to account for these observations. The formation of the *endo* product was suggested to have a lower activation energy barrier and therefore to be the preferred pathway at low temperature and when the reaction was left for short period of time. On the other hand, the *exo* product was suggested to be thermodynamically favourable, and therefore predominant when products reach equilibrium (Fig 2). Although the terminology of thermodynamic/kinetic control was not used in this original work, and would only begin to appear in the literature years later, Woodward provided the first implicit use of these concepts in the field of organic chemistry. The concepts were later formally defined by Catchpole, Hughes, and Ingold, in 1948, who emphasised the distinctions between kinetic and thermodynamic control in an allylic rearrangement process, using a similar graphical illustration to that originally presented by Woodward and Baer. Since then, these concepts have become an essential part of the chemical lexicon in rationalising reactions and selectivity.



**Scheme 2** The computed energy profile (SMD-M06-2X/def2-QZVPP) shows the competition between thermodynamic and kinetic control as exemplified by Woodward and Baer for the diastereoselective cycloaddition of 6,6-pentamethylenefulvene with maleic anhydride<sup>3</sup>.

The change between a kinetically to a thermodynamically controlled process can be influenced by temperature, but also by the identity of reagents, catalysts, and/or the solvent used. Additionally, while, it is possible that diastereoselectivity may result in either scenario, enantioselectivity is restricted only to reactions under kinetic control, since enantiomers are isoenergetic. Although it is instructive to consider the single-step reactions in Fig. 2A and Fig. 2B, as graphical examples of thermodynamically and kinetically controlled processes, respectively, most synthetic transformations occur in a multistep fashion. One such scenario is envisaged in Fig. 2C, where the irreversible product forming step is preceded by the interconversion of reactant conformations or diastereomers. Provided the intermediates can rapidly equilibrate, the Curtin-Hammett principle enables the quantitative estimation of product stereoselectivity. However, for reactions involving several competitive pathways one may require the use of more complex kinetic models in order to account for stereoselectivity.

The next sections emphasise the application of these concepts in generating quantitative estimates of selectivity from computations, reliant upon the underlying mathematic relationships.



**Fig. 2** Illustrative Gibbs energy profiles for single-step reactions under (A) thermodynamic and (B) kinetic control. In the first case the products reach equilibrium, while in the latter conversion is irreversible. (C) Application of the Curtin-Hammett principle when two species  $R_A$  and  $R_B$  equilibrate more quickly than the forward reaction. In this particular example, the minor intermediate affords major product A.

## 2.1 Stereoselectivity under thermodynamic control

To quantitatively illustrate the concepts introduced above, we will consider a single-step process such as that illustrated by the Gibbs energy profile in Fig. 2. Here, a single reactant R can be converted into two stereoisomeric products A and B. Firstly we discuss the scenario where interconversion between these products is possible, either through reversibility of the first step or by an additional pathway. If the relative amounts of A and B are considered to be time-independent (*i.e.* either because they have reached equilibrium, or the competing pathways are of the same order), the product ratio  $[A]/[B]$  is determined by their relative stabilities. The reaction is said to be *thermodynamically controlled* (Fig. 2A). Stereoselectivity is mathematically expressed in terms of the temperature-dependent equilibrium constant, which is related to the standard-state Gibbs free energy of reaction,  $\Delta G^0$  at temperature T:

$$\frac{[A]}{[B]} = K = e^{-\Delta G^0/RT} \quad (1)$$

where  $K$  represents the equilibrium constant and  $R$  is the gas constant. Since the relative Gibbs energies of enantiomers are identical (*i.e.*  $K = 1$ ) it is obvious that enantioselectivity is impossible under such conditions. Diastereoselectivity is a possibility, its extent dependent upon the relative stability of diastereomers A and B.

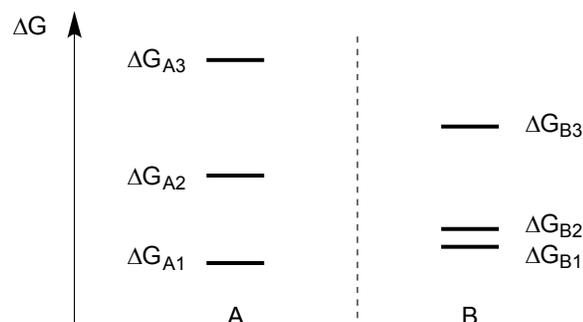
Thermodynamic stabilities can be determined experimentally by calorimetric measurements or studies of equilibrium constants *e.g.* by NMR. Computationally, the relative stability of two diastereomers can be computed directly from the difference in calculated Gibbs energies at the appropriate temperature (concentration/pressure do not affect the diastereomeric  $\Delta G_0$ ). Often this may be approximated to the difference in electronic energies, effectively assuming that the partition functions of the two species are identical. Diastereomeric energy differences obtained from fairly modest calculations using density functional theory (DFT) with a medium-sized basis are often within 1–2 kcal·mol<sup>-1</sup> of experiment, despite worse performance for activation barriers and reaction energies. Most species will exist as several, rapidly interconverting, conformations that contribute to the equilibrium ratio. Computationally it is necessary to compute the stability of each of these conformers individually (Fig. 3).

In these more complex cases, where different conformations of the same product are accessible at the experimental temperature, one can calculate the ratio of the population of products as a function of their Gibbs energy differences and the Boltzmann factors associated to each of them in Eq 2. Here, the numerator sums all Boltzmann factors associated with conformers of product A and the denominator sums all those associated with conformers of product B (also see SI)

$$\frac{[A]}{[B]} = \frac{1 + e^{(\Delta G_{A1} - \Delta G_{A2})/RT} + e^{(\Delta G_{A1} - \Delta G_{A3})/RT}}{e^{(\Delta G_{B1} - \Delta G_{B2})/RT} + e^{(\Delta G_{B1} - \Delta G_{B3})/RT}} \quad (2)$$

where  $\Delta G_{xi}^x - \Delta G_{yj}^x$  is the Gibbs energy difference between the product (X,Y = A,B,...) in their different conformations ( $i,j = 1,2,\dots$ ) and A1 is the most stable. In the scenario depicted above, A would be the major diastereomer. However, considering only the stabilities of the most stable species would overestimate the

selectivity since the second and third conformers of B are more stable than those for A.



**Fig. 3** Scenario in which diastereomeric products A and B are characterised by three thermally accessible conformations.

## 2.2 Stereoselectivity under kinetic control

If the activation barriers for the conversion of either A or B back into R (see Fig 2) are large enough to prohibit the reverse reaction and there is no product interconversion, the reaction will be *kinetically controlled*. Any stereoselectivity will be determined by the relative rates of formation of each product (Fig. 2B).

Within the context of transition-state theory (TST), the macroscopic rate constant for the formation of each product can be expressed as in Eq 3:

$$k_1 = \frac{k_B T}{h} e^{-\Delta G_A^\ddagger/RT} \quad k_2 = \frac{k_B T}{h} e^{-\Delta G_B^\ddagger/RT} \quad (3)$$

where  $\Delta G_{A/B}^\ddagger$  are the activation free energies for formation of A or B, respectively,  $k_B$  is the Boltzmann constant, and  $h$  is Planck's constant. A similar treatment to the one shown above (Eq 2) can be used in case several conformers are accessible at the reactant and transition state. The application of TST has proven sufficient to quantitatively account for the rates and selectivities of many thermal reactions. However, the expressions in Eq 3 only holds under certain assumptions<sup>4</sup> and recent works have revealed instances where TST is insufficient to describe common organic reactions<sup>5,6</sup>.

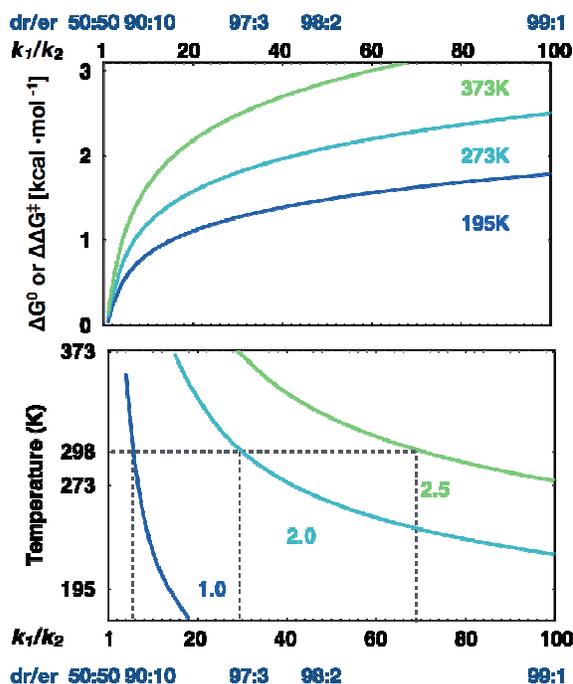
A key assumption of TST is that redistribution of internal (vibrational) energy within and between molecules/solvent is faster than the timescale for breaking/forming a bond. This means that reactants and the activated complex are close to thermodynamic equilibrium (*quasi-equilibrium*) and the rate of the process follows Maxwell-Boltzmann statistics. Within this framework the reaction is also assumed to be electronically adiabatic, *i.e.* the Born–Oppenheimer separation of electronic motion from internuclear motions is valid. Furthermore, the TS is described as dividing surface perpendicular to the reaction coordinate and all trajectories passing through this surface go on to product without recrossing. This means that the rates calculated from Eq 3 will always provide an upper limit to the true rate constant<sup>4</sup>. To allow for recrossing, a temperature-dependent *transmission* coefficient  $\kappa$  (usually close to 1) can be introduced, which also allows one to account for quantum effects. Finally, within the TST framework, it is assumed that at all times during a chemical reaction the system follows the minimum energy path (MEP) on the potential energy

surface (PES), thus neglecting the kinetic energy of the system. Even though this can be considered a fairly good approximation, several examples have been shown in recent years emphasising the importance of dynamic factors in dictating the chemical output. This is the case for example, when a very shallow intermediate exist or where a valley bifurcates into two (so-called *bifurcation points*)<sup>5</sup>. In these cases, a single TS lead to two different products, and therefore selectivity is no longer determined at the TS.

Keeping in mind the limits of its predictive power, we outline here the principles of TST as it is used to computationally model selectivity. Within this framework and at a given temperature, selectivity in a kinetically controlled reaction will be given by the ratio between the competitive rates (Eq 4).

$$\frac{[A]}{[B]} = \frac{k_1}{k_2} = e^{-\Delta\Delta G^\ddagger/RT} \quad ; \quad \Delta\Delta G^\ddagger = \Delta G_A^\ddagger - \Delta G_B^\ddagger \quad (4)$$

Under mild conditions kinetic control is the most common scenario. Moreover, for enantioselective reactions, where products are isoenergetic, this is the only possibility, unless a chiral solvent or resolving agent is used, in which case diastereomeric adducts with different thermodynamic stabilities can be formed. Enantioselective induction can also be obtained using chiral catalysts (asymmetric catalysis). In this case a chiral catalyst selectively interacts with a prochiral substrate to generate a pair of diastereomeric transition states with different energies, which evolve to enantiomeric products. The enantiomeric product that predominates is the one generated *via* the TS lower in energy, despite the fact that both products have identical energies.



**Fig. 4** Dependence of kinetically controlled stereoselectivity ( $k_1/k_2$ ) on temperature (top panel); and free energy difference between the TS/products for a kinetic/thermodynamically-controlled reaction (bottom panel).

### 2.3. Quantifying selectivity

In either a kinetically or thermodynamically controlled regimen, it is important to emphasise the exponential dependence of the product distribution on both free energy and temperature. This is graphically illustrated in Fig. 4, where Eqs 1 and 4 are plotted for different temperature and free energy differences. As can be seen, only 1 kcal·mol<sup>-1</sup> difference between the TS/product energies is enough to obtain a product ratio of 90% or more. Additionally, at low temperature, where the curves flatten out more quickly, such dependence becomes even steeper. This means that for values of  $\Delta\Delta G^\ddagger$  a precision in the order 1.0 kcal·mol<sup>-1</sup> or less is necessary to obtain reliable estimates and draw meaningful conclusions.

As shown in Fig. 4, stereoselectivity can be expressed either as a diastereomeric (dr) or enantiomeric ratio (er) or as percentage of the major product, which are directly related to the Gibbs energy difference between TSs or products. Given the historical relationship between chirality and optical activity, another way to quantify the product composition is optical purity ( $O_p$ ), which is defined in terms of specific rotations measured by polarimetry:

$$O_p = 100 \times ([\alpha]_{\text{mixture}} / [\alpha]_{\text{pure sample}}) \quad (5)$$

where  $[\alpha]_{\text{mixture}}$  and  $[\alpha]_{\text{pure-sample}}$  are the observed and maximum specific rotations of a sample. Assuming a linear relationship between optical activity and composition,  $O_p$  is numerically equivalent to the percentage of one enantiomer over the other, which is defined as *enantiomeric excess* (ee):

$$ee = \frac{[R] - [S]}{[R] + [S]} \times 100 \quad (6)$$

where  $[R]$  and  $[S]$  represent the mole fraction of the *R* and *S* enantiomer, so that a mixture composed of 80% *R* enantiomer and 20% *S* enantiomer will have an ee of 80% – 20% = 60%. Despite the fact that today few chemists use  $O_p$  to measure optical purity and the merits of alternative metrics have been suggested<sup>7</sup>, ee has become the most common way to report enantioselectivity. For mixtures of diastereomers, analogous terms for diastereomeric excess and percentage diastereomeric excess have also been defined.

From Eqs 4 and 5, one can computationally determine ee as in Eq 7. As with ground-state structures, the absolute Gibbs energies are obtained for the pair of diastereomeric TSs leading to the enantiomeric products.

$$ee = \frac{1 - e^{-\Delta\Delta G_{R/S}^\ddagger/RT}}{1 + e^{-\Delta\Delta G_{R/S}^\ddagger/RT}} \times 100 \quad (7)$$

In a *Curtin-Hammett* scenario (Fig. 2C), selectivity can be estimated by computing the standard free energy difference between the respective transition states only (as in Eq 4 and/or 7). Here, the information regarding the energy difference between the intermediates is considered to be irrelevant. However, it is important to keep in mind that this is implicitly contained in the activation term ( $\Delta\Delta G^\ddagger = \Delta G_A^\ddagger - (\Delta G_B^\ddagger + \Delta G_{AB}^\circ)$ ). This treatment can be extended to any case where different products are formed from two rapidly interconverting starting materials (conformers, tautomers or isomers). A well-characterised example of such a situation is the asymmetric hydrogenation of  $\alpha$ -

acetamidocinnamates with a chiral rhodium-phosphine catalyst proceeding via the Halpern-Brown mechanism, in which the minor (but more reactive intermediate) undergoes reaction to afford the major enantiomer of product<sup>8</sup>.

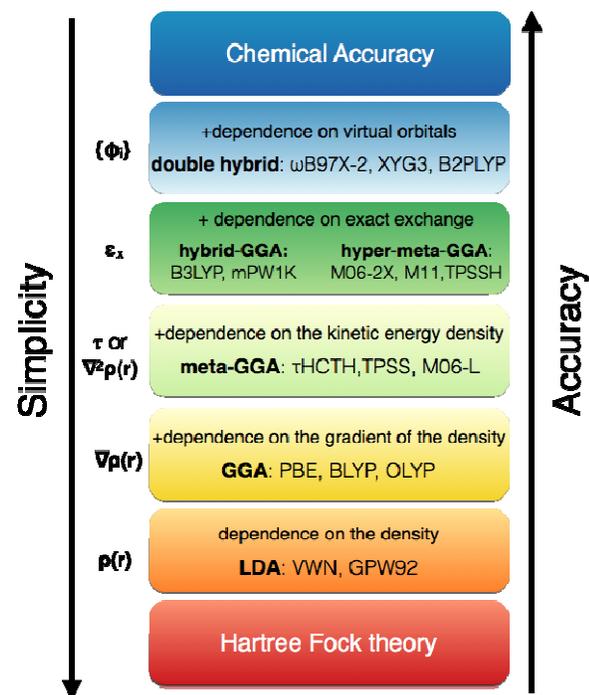
### 3. Theoretical Methods for Computing Stereoselectivities

The ability to model structures and energetics of competing TS structures underlies much computational effort in exploring the origins of stereoselectivity. In general terms, quantum mechanics (QM) calculations are most commonly employed for this task since bond formation and/or breaking requires no additional parameterization, unlike classical molecular mechanics (MM) methods. This is not to say that MM has no place in the study of stereoselectivity, since it can be useful for conformational sampling prior to QM computations, as a part of hybrid QM:MM methods to describe larger chemical systems, or in specific MM parameterisations to describe TS structures<sup>9</sup>. Houk has schematically described a workflow, from mechanistic hypothesis, progressing through MM conformational searches to QM TS optimisations and evaluation<sup>10</sup>. Nonetheless, chemical systems on the order of hundreds of atoms are currently amenable to study with QM methods, and in particular density functional theory (DFT) has become a staple in the study of reaction mechanisms and selectivities. DFT offers the advantage of describing instantaneous electron-electron correlation, which is neglected in the mean field Hartree-Fock (HF) approach, at a much lower computational cost compared with correlated wave function theory (WFT, often referred to as *ab initio*). For example, while HF and DFT methods scale nominally as  $N^3$ - $N^4$  ( $N$  being a relative measure of the system size), correlated WFT approaches such as Moller-Plesset (MP2) and Coupled-Cluster (CC) scale as  $N^4$  and  $N^6$ - $N^7$ , respectively. Although HF calculations were key in seminal early studies of diastereoselectivity<sup>11</sup>, today they have been largely superseded by DFT. Even though correlated WFT calculations are now possible, they suffer from slower convergence with respect to basis-set size, which has contributed to the establishment of DFT as the “method of choice” for the study of realistic chemical reactions. Recent methodological developments in local forms of correlated *ab initio* calculations have enabled them to be applied to the study of catalytic reactions; however, geometry optimisations remain out of reach at present<sup>12</sup>.

Today, a large number of DFT exchange-correlation functionals exist (known as “the functional zoo”)<sup>13</sup>. While this offers flexibility in terms of choice, it is also challenging to systematically explore the different features, advantages, and pitfalls inherent to these different methods. Since the exact form of the universal density functional is unknown, each functional adopts a different formulation of inter-electronic exchange and correlation with varying levels of empiricism. Some only use parameters based on general rules of quantum mechanics (LDA, PBE, TPSS, and TPSSH), while other use a variable number of empirical parameters fitted to experimental data (B88, B3LYP,  $\omega$ B97X-D, M06-family, etc). Furthermore, they can also be classified according to their rung on Perdew's ‘Jacob's Ladder’, each rung representing a different level of approximation of the exchange-correlation functional (Fig. 5).

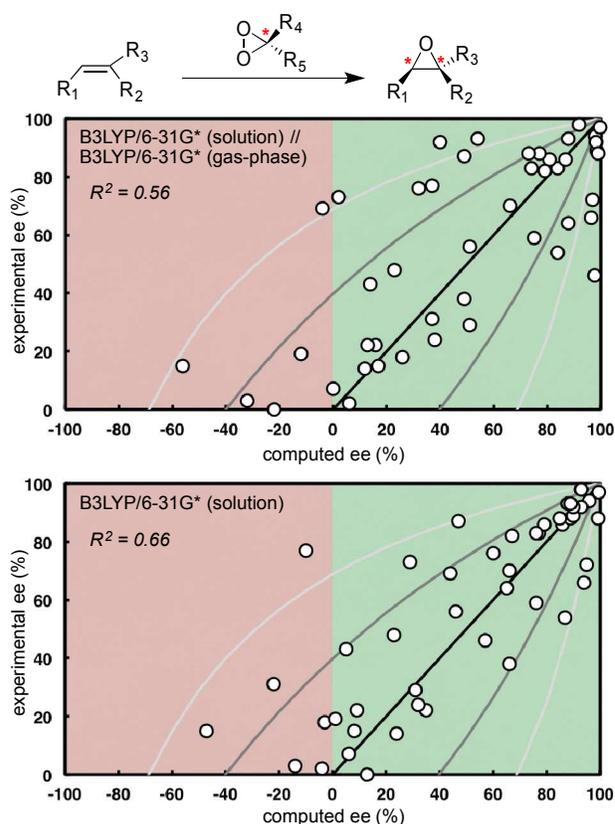
In general, there is no single functional that can perform best for all chemical applications and therefore a careful choice needs to be made for each system under study. There is an abundance of DFT functional benchmarking studies for chemical applications to

be found in the literature, and while there are relatively few studies specific to modelling stereoselectivity<sup>14-17</sup>, nonetheless we attempt to summarise the important considerations here:



**Fig. 5** The hierarchy of exchange-correlation functionals is often represented by the rungs of *Jacob's ladder*, according to the number and kind of local ingredients.

There are well-publicised failures of the B3LYP hybrid functional to capture the effects of dispersion (*i.e.* medium or long-range correlation effects) or to describe energy changes accurately in simple reaction types<sup>18</sup>. Despite this, the B3LYP/6-31G(d) (equivalently indicated by 6-31G\*) combination of functional and basis set has been applied to the study of stereoselective reactions to optimise competing TSs and to evaluate their energies, giving satisfactory agreement with experiment<sup>10</sup>. Houk's seminal work in uncovering the mechanisms and selectivities of organocatalytic aldol and Mannich reactions (discussed in more depth below) adopted this level of theory and sufficient quantitative accuracy was possible to enable the design of new stereoselective reactions, a landmark in the field<sup>19</sup>. Rzepa's comprehensive analysis of Houk's earlier work, demonstrated that the resulting Houk-List model of stereoselection, remains robust irrespective of the computational treatment adopted<sup>17</sup>. Corrections for dispersion caused a general decrease in energy differences between stereoisomeric TSs, but the impact on the predicted stereoselectivities was nearly negligible for the systems originally studied by Houk and List<sup>13</sup>. Simon and Goodman have systematically studied the choice of functional on the gas phase energy difference between two competing TS structures for nineteen organic/organometallic reactions<sup>14</sup>. Their work revealed that B3LYP optimizations coupled with re-evaluation of the energies (single-point calculations) at these geometries using more recently developed functionals (in this case M05-2X) were appropriate for many studies of organic reaction mechanisms.



**Fig. 6** Comparisons of B3LYP/6-31G(d) computed enantioselectivities against experimental results in asymmetric epoxidation of alkenes by dioxiranes from *ref.*<sup>15</sup>. The green area represents reactions for which the absolute *sense* of enantioselectivity was successfully predicted, whereas the red area represents a prediction in the opposite sense to experiment.

The absence of “experimental” TS geometries means that experimental selectivities, kinetic isotope effects, or high-level computed structures/energetics are used as computational benchmarks for optimised TS structures. Tsogoeva and Clark focussed on stereoselective organocatalytic transformations and compared the gas phase TS geometries obtained from DFT calculations with those from more expensive MP2/TZVP optimisations<sup>16</sup>. In this study, more modern functionals which incorporate an explicit or implicit correction for dispersion ( $\omega$ B97X-D and M06-2X) gave structures closer to the *ab initio* benchmark, and energetics were closer to the reference values using the M06-2X method<sup>20</sup>.

The most comprehensive “meta-analysis” of the ability to quantitatively predict enantioselectivity in a given asymmetric transformation with DFT calculations has been performed by Breslow and Friesner, who have studied the epoxidation of alkenes by chiral dioxirane reagents<sup>15</sup>. Such studies are relatively rare since there is often limited experimental selectivity data for a given reaction, coupled with the fact that the computational analysis of conformationally flexible systems is still far from routine. Nevertheless, this study demonstrates the value of recording and utilising all selectivity values (which often remain unpublished) and also the fairly good performance achieved with B3LYP/6-31G(d) computations. In **Fig. 6** we show the comparison of computed and experimental selectivities obtained for enantioselective

epoxidation<sup>15</sup>, obtained from the solution-corrected TS free energy differences, as in Eq 7.

The two panels in Fig. 6 result from gas phase and solution-phase optimisations. Most of the time, the computed sense of enantioselectivity agrees with experiment (green section), with relatively few outliers as qualitatively incorrect (red section). The straight line represents perfect quantitative prediction, while the arcs around this line encompass regions that lie within  $\pm 0.5$  kcal·mol<sup>-1</sup> (dark grey) and  $\pm 1.0$  kcal·mol<sup>-1</sup> (light grey) from the experimental free energy difference. TS optimisations in solution give only one result which is clearly outside the larger error-bound. The authors note that the correlation coefficient is high enough that it would be plausible to use this computational methodology to predict low, medium or high levels of enantioselectivity. It is also interesting to note that the quality of these predictions were worsened using energetics obtained with the M05-2X functional. This result emphasises the importance of benchmarking a chosen level of theory or density functional carefully against empirical results before attempts at prediction are made.

From an empirical perspective, the use of DFT optimisations with fairly modest double-zeta valence polarized basis sets, such as B3LYP/6-31G(d) has been, and continues to be a cost-effective approach for the study of kinetically-controlled stereoselectivities. The comparison of chemically-related structures, as in a stereoselective reaction, benefits from the cancellation of systematic errors, and quantitative accuracies in several cases compare well with experiment. Equally, systems with interactions where B3LYP may be expected to fail, where sizable dispersive interactions are important, will require the inclusion of dispersion corrections either by design (the Minnesota functionals) or from its explicit inclusion (DFT-D, vdW-DF). Another alternative to include long-range interactions is to use an atom-centered potentials, so called dispersion-correcting potentials (DCPs)<sup>21</sup>. QM/MM calculations may offer the best of both worlds, combining a QM-description of changes in chemical bonding with the explicit van der Waals terms of the classical approach at a reduced cost compared to a semi-empirical or quantum chemical lower level, which lack an adequate description of dispersion terms<sup>22</sup>. We note that the effects of truncated basis sets tend to mimic attractive dispersion forces (albeit without the correct R<sup>-6</sup> dependency at long range), such that optimisations with small basis sets and dispersion corrections will give misleading results. Recent experimental results also suggest that the extent of measured dispersion interactions in solution are exaggerated by DFT-D, presumably by the neglect of solute-solvent interactions in the gas phase calculations<sup>23</sup>.

## 4. Computational Studies of Asymmetric Catalysis: Case Studies

### 4.1 Proline-catalysed asymmetric aldol reactions

Asymmetric organocatalysis emerged from the seminal work of Hajos, Parrish, Wiechert, Eder and Sauer in the early 1970s using proline as a catalyst. More recent developments have been pioneered by List, Barbas and Lerner<sup>24</sup> and Macmillan<sup>25</sup>. This has led to the generation of new organocatalysts, including Lewis acids and bases, Brønsted acids and bases, and chiral phase transfer catalysts<sup>25, 26</sup>. In parallel with these synthetic developments, computational studies have provided simple yet powerful models to rationalise the origin of enantioselectivity in these processes<sup>2, 10</sup>. Here, we briefly describe some of these works, and show how different computational tools have been used in order to probe the

mechanism and the origin of stereoselectivity of challenging chemical systems.

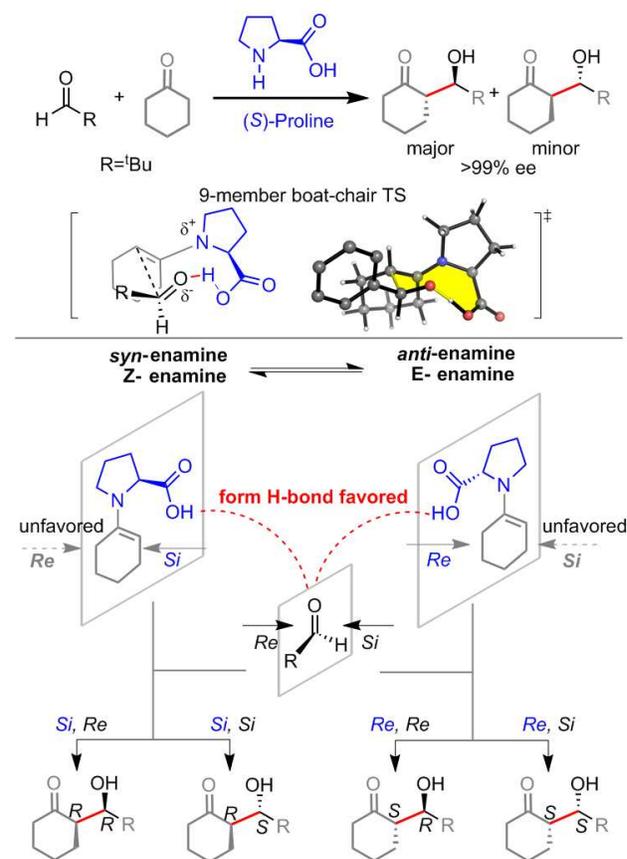


Fig. 7 Stereochemical possibilities for the asymmetric aldol reaction based on Houk-List model.

In 2003 Houk, List and colleagues investigated the proline-catalysed asymmetric intermolecular aldol reaction between cyclohexanone and benzaldehyde (and isopropionaldehyde)<sup>19</sup>. The computational procedure used in this study, as well as the model derived from it (known as the *Houk-List model*, Fig. 7) provided a generally applicable transition state model for C-C bond formation, as well as the basis for rationalising the sense of the observed stereoselectivity in this and related systems.

This investigation began by examining the potential products and transition states of the reaction. Considering the relative orientation between the enamine and the acid, *anti* or *syn*, and the two prochiral faces for attack, *Re* or *Si* (Fig. 7), they found four different stereochemical products. Additionally, considering the three different staggered arrangements between aldehyde and enamine and both half-chair conformations of the cyclohexene ring, the authors found 24 potential TSs. Using model systems to analyse key interactions they were able to reduce this number to eight relevant TSs (for four possible stereoisomeric products). This preliminary exploration substantially reduced the conformational space to be explored, and consequently the number of calculations that had to be performed. These calculations included geometry optimisation of stationary points (reactants and TSs) in gas phase at the B3LYP/6-31G\* level of theory, followed by single-point

calculations using a polarizable continuum model (CPCM) of DMSO solvent.

TSs involving *re-face* attack of the *anti*-enamine were found to be the most favourable ones (Fig. 7). These TSs were characterised by having a half-chair conformation of the NCC–CO atoms and a hydrogen bond interaction between the carboxylate and the aldehyde, which was found to play a key role in determining the outcome of the reaction, (Fig. 7, middle panel). Furthermore, a hydrogen bond interaction between the hydrogen of the carbon adjacent to the proline nitrogen and the forming alkoxide (NCH...O) was found to provide an additional electrostatic stabilisation to these conformations. This model of stereoselection agreed with the absolute sense of stereoselectivity, showing that the (*S,R*) diastereomer is the dominant species formed, and the quantitative levels of selectivity observed experimentally. The pyrrolidine ring of the proline catalyst is non-planar, and its conformation (up/down) can influence stereoselectivity depending on the substituents, because stabilisation of the forming alkoxide changes as the proline ring flips between these two conformations. These and other more subtle conformational changes show the importance of a thorough search of conformations of transition states for accurate predictions of stereoselectivities<sup>27</sup>.

Recently Rzepa and colleagues<sup>17</sup> have revisited this reaction and examined the effect of using modern computational methodologies on the final results, including the use of a larger basis set, inclusion of implicit solvent, dispersion, kinetic isotope effects (KIE) calculations, and analysis of non-covalent interactions (NCI). This study demonstrated the robustness of the Houk-List model, confirming the role of the NCH...O interaction and the nature of the rate-limiting step. The alternative mechanisms studied (such as proton-relays involving a water molecule or the Hajos–Parrish alternative) were found to be higher in energy. This demonstrates that despite the quantitative improvements of current methodologies, the computational protocol originally presented by Houk and List and the model generated remain valid today.

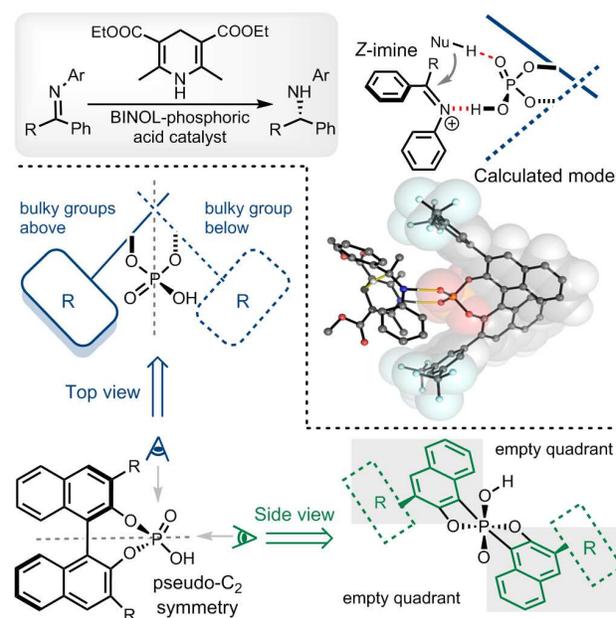
## 4.2 BINOL-phosphoric acid catalysis

In addition to the covalent mode of activation exhibited by proline and imidazolidinone-based organocatalysts, chiral organic molecules can also induce stereoselectivity through non-covalent interactions. Using this activation mode, a vast number of highly enantioselective organocatalysts have been developed over the last decade, including ureas, thioureas and BINOL-derived phosphoric acids<sup>28</sup>. This area has also proved to be fruitful for the computational understanding of catalytic asymmetric synthesis, where the quantitative results from calculations led to the establishment of simplified structural models that have found use in predicting selectivities for a diverse array of reactions.

Such synergy can be exemplified by the computational studies on the catalytic mechanism of BINOL-derived phosphoric acids, originally developed by Akiyama and Terada<sup>29, 30</sup>. Símon and Goodman<sup>31, 32</sup> and Marcelli, Hammar and Himo<sup>33</sup> independently studied the asymmetric reduction of imines by Hantzsch esters (dihydropyridines) catalysed by C<sub>2</sub>-chiral phosphoric acids (Fig. 8). Due to the large size of the real catalysts, Símon and Goodman<sup>31</sup>, used QM:MM calculations to optimise transition structures of the full catalyst at the ONIOM(B3LYP/6-31G(d):UFF) level of theory. The Universal Force Field (UFF) description of conformational energetics is typically quite poor<sup>34</sup>, however, when used to study a rigid catalyst backbone and flanking groups this is not as important as

the force field's description of non-bonding interactions between the catalyst's aromatic groups and the two substrates. This can lead to greater accuracy than uncorrected DFT (e.g. B3LYP) calculations<sup>31</sup>. Marcelli, Hammar and Himo used a full DFT description<sup>33</sup>, optimising at the B3LYP/6-31G(d) level, to describe the resolution of chiral  $\alpha$ -branched aliphatic imines, which necessitated the truncation of the binaphthyl to a biphenyl backbone in the catalyst.

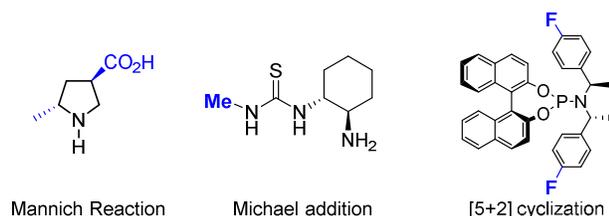
Both of these computational studies established a bifunctional role for the phosphoric catalyst. The acidic proton activates the imine substrate through a hydrogen bond, while the P=O bond simultaneously coordinates to the N-H bond of the reducing agent (Fig. 8). With both substrates coordinated to the phosphoric acid group in the reduction TS, the 3,3'-substituents of the binaphthyl backbone create a C<sub>2</sub>-symmetric chiral environment that discriminates between the attack of the two enantiofaces (diastereofaces in the resolution) of the imine. The energetically favoured TS proceeds *via* a Z-imine geometry in which the N-substituent is oriented into an open region of space. The computed structures are encapsulated by the working models of Goodman, Himo and Terada<sup>35</sup> which have subsequently been deployed to understand and predict selectivities in the reactions of C<sub>2</sub>-chiral phosphoric acids with several different substrates<sup>32</sup> and related bis-iminophosphoranes. Reaction types with this class of catalyst studied computationally now include (in addition to reaction of an imine derivative with nucleophiles), Michael additions, Mannich-type reactions, Friedel-Crafts Reactions, Dipolar cycloadditions, among others (see ref. <sup>26</sup> and references cited therein).



**Fig. 8** Enantioselectivity in the transfer hydrogenation of aryl-imines. Concise model of BINOL-phosphoric acid used by Goodman<sup>31</sup> (top view), and Himo<sup>33</sup>, Terada<sup>35</sup> (side view). Here, the dihydropyridine reagent was modeled as the methyl diester.

## 4.3 Enantioselective catalysts by computational design

Considerable synthetic effort is currently invested in the discovery and refinement of new asymmetric catalysts. Traditionally, trial-and-error approaches have been used, with mechanistic insights

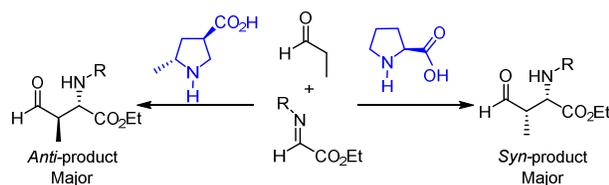


**Scheme 3** Asymmetric catalysts designed by computation and validated experimentally. Blue substituents indicate modifications from existing structures screened computationally.

following rather than leading their synthesis. However, with the development of field, some few notable examples have shown the great potential of computational modelling to guide asymmetric catalyst design<sup>36–38</sup> (**Scheme 3**). Thus far, this work has focused on computational-led modification of existing catalyst or ligand scaffolds. While the *de novo* computational discovery of new asymmetric catalysts remains out of reach, it has proven possible to design ligands that make reactivity possible which would not otherwise occur<sup>39</sup>.

#### 4.3.1 Anti-diastereoselectivity in the Mannich reaction by a designer organocatalyst

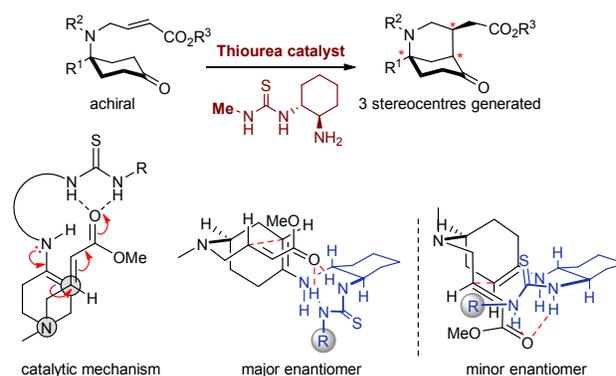
The first and successful examples of the computational design of a stereoselective organocatalyst that was later experimentally validated resulted from collaborative work between Barbas and Houk groups (**Scheme 4**). They focused on the development of an unnatural amino acid to catalyse Mannich reactions giving opposite diastereoselectivities to proline<sup>36</sup>. The change from *syn*- to *anti*-selectivity was engineered by fixing the enamine conformation. In the newly designed organocatalyst a methyl group was introduced to the 5-position and the carboxylic acid was moved around to the 3-position, with the two pyrrolidine substituents *trans*- to each other. Similar to proline, imine facial selectivity is controlled by proton-transfer from the acid to the imine, whereas steric interactions due to the methyl group force the enamine to react from the opposite enantioface. Computationally, the four competing TS stabilities (*i.e.* according to the Curtin-Hammett principle) for the addition of propionaldehyde to *N*-PMP-protected  $\alpha$ -imino methyl glyoxylate were considered at the relatively cheap HF/6-31G(d) level of theory. This catalyst was predicted to give the desired inverted 95:5 *anti*:*syn* diastereoselectivity and 98% ee for the formation of the (2*S*,3*R*)-product. Experimentally, the addition of propionaldehyde and ethyl glyoxylate gave a 94:6 dr and > 99% ee, with the same sense as predicted computationally.



**Scheme 4** A computationally designed catalyst for the asymmetric Mannich reaction gives access to the anti-diastereomer, in contrast to the *syn* selectivity obtained with proline.

#### 4.3.2 A computationally modified primary amine for enantioselective Michael additions

Together with Prof. Darren Dixon (Oxford) we have recently studied the organocatalysed intramolecular Michael addition of ketones to  $\alpha,\beta$ -unsaturated esters<sup>37</sup>. This reaction is promoted by primary amines, forming only one diastereomer in the process. The development of this transformation into an enantioselective version by using chiral amines was studied in tandem by experiment and computation. Computationally, a novel catalyst structure was examined and predicted to be highly enantioselective: this was later confirmed synthetically (**Fig. 9**). Firstly, the mechanism of the reaction was investigated using M06-2X/6-311+G(d,p) computations. The use of a triple-zeta valence polarized basis set and the inclusion of solvation (toluene) during the geometry optimisations was found to be important to describe the stepwise nature of the C-C bond forming step. The *endo*-diastereoselectivity of this reaction was found to result from a lower barrier and irreversible nature of the Michael addition step.



**Fig. 9** Computed mechanism of primary-aminothiourea promoted intramolecular Michael addition, showing the computed TS geometries for the major and minor enantiomer, which led to the computationally proposed modification of the R-substituent.

For the asymmetric reaction, the different possibilities were computed separately: the enamine and enoate can react via either enantioface and the enamine can adopt either an *s-cis* or *s-trans* conformation. The conformation of a bifunctional aminothiourea catalyst was investigated by fixing the core TS structure and sampling many possible structures with semi-empirical (dispersion-corrected PM6) calculations. This method, an abbreviated form of HF calculations by approximating or neglecting integrals, gives results qualitatively consistent with DFT calculations and reasonably accurate optimised geometries, making it useful for preliminary studies. Considering enantioselectivity, the reactive enamine conformation differs between the two pathways, with the *s-cis* enamine yielding the major enantiomer and the *s-trans* enamine yielding the minor enantiomer. The conformation of the catalyst also differs between the two enantiomers, with eclipsing interactions in the minor pathway. Although all these conformations benefit from close NH...O contacts between the ester and catalyst, our computational study suggested a lack of any significant contribution from the thiourea substituent (R in **Fig. 9**).

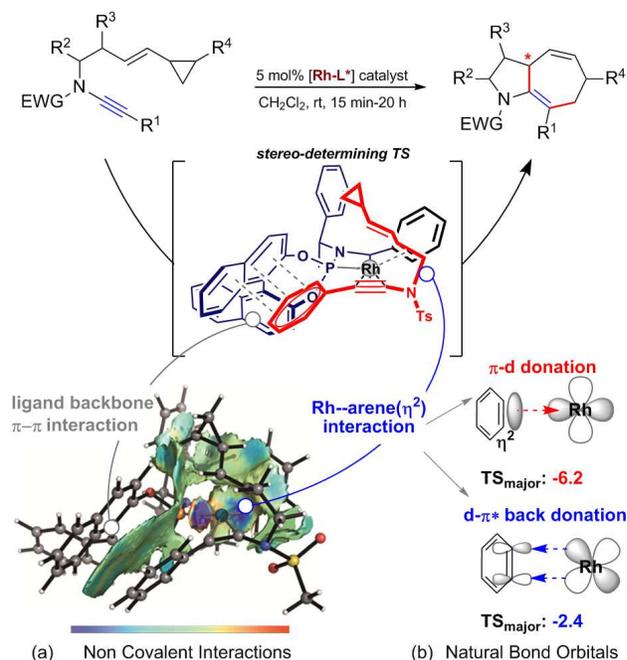
Therefore, we proposed that the catalyst could in principle be simplified to –Me from a more complex structure. Accordingly, this new catalyst was synthesised and tested. The product of the reaction was obtained in 83% yield and 97% ee as a single diastereomer, thus validating the computational hypothesis. This work illustrates how a detailed analysis of the mechanism and the key interactions driving stereocontrol can inform design of a simpler and robust catalyst.

#### 4.3.3 Ligand optimisation in asymmetric transition-metal catalysed cycloisomerisation

It is possible to quantitatively account for levels of stereoselectivity in terms of the relative energy difference between diastereomeric transition state structures that define the competitive pathways. However, the rational modulation or design of new catalysts also requires an understanding of the structural or electronic features underlying this selectivity. Currently, several computational techniques are available to study such effects. They include topological approaches, such as Bader's atoms in molecules (AIM) theory and noncovalent interaction (NCI) analysis, which utilise the electron density, and orbital-based approaches, such as the natural bond orbital (NBO) method.

A collaborative experimental and computational investigation with Prof. Edward Anderson (Oxford) led to the discovery of new chiral phosphoramidite ligands for use in Rh-catalysed stereoselective [5+2]-cycloisomerisations of ynamide-vinylcyclopropanes (Fig. 10)<sup>38</sup>. Firstly, we compared competing mechanisms for the cycloisomerisation mechanism, at the  $\omega$ B97X-D/6-311++G(d,p)/LanI2TZ// $\omega$ B97X-D/6-31G(d)/LanI2DZ level of theory. A dispersion-corrected functional was chosen due to the presence of nonbonding interactions, evident in the X-ray structures of metal-phosphoramidite complexes. All stereochemical possibilities were computed, including the substrate and ligand orientations about the coordinating metal centre. These computations suggested a mechanistic sequence in which an irreversible, stereodetermining C-C coupling between the alkyne and alkene takes place before metal insertion into the vinylcyclopropane (Fig. 10). The enantioselectivity computed according to this model for Feringa's original phosphoramidite ligand (in which F atoms are replaced by H, Scheme 3) matched experimental data (calculated  $\Delta\Delta G^\ddagger$ : 2.7 kcal·mol<sup>-1</sup>, 97.9% ee comparing with 98% ee in the experiment), and so we sought to understand how chemical modification of the phosphoramidite ligand could be used to enhance stereoinduction.

Intramolecular interactions in the key TS were analysed separately with both NCI and NBO approaches (Fig. 10). Both are available as standalone programs, while NBO v3 is implemented as a fully integrated part of the Gaussian package. The NCI isosurface shows regions in space corresponding to overlapping atomic densities, which may be associated with noncovalent interactions. The coloration is used to indicate the approximate magnitude and sign of these interactions. In this case, green regions showing the van der Waals/dispersion-dominated interaction between ligand backbone and substrate, and blue regions showing a strong attraction between Rh and the P atom and aromatic group of the ligand. Red regions, indicative of steric repulsion are absent in this analysis. NBO analysis allowed us to quantify specific donor-acceptor interactions between the metal and the aromatic group in this structure. The major contribution comes from a  $\pi$ -donor orbital of the arene into the vacant d-orbital of Rh, rather than due to Rh back donation to  $\pi^*$  of C=C (Fig. 10b).



**Fig. 10.** Non-Covalent Interaction (NCI) and Natural Bond Orbital (NBO) analyses of key interactions involving the phosphoramidite ligand during the asymmetric [5+2] cycloisomerisation process.

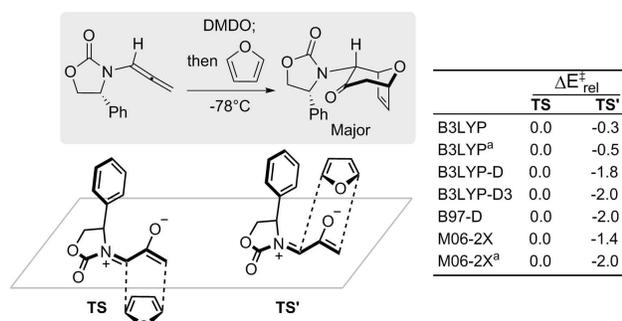
Identifying an arene-metal interaction in the stereodetermining TS prompted attempts to strengthen or weaken this interaction. As expected from the NCI and NBO analyses, phenyl substitution by an electron-donating group (p-OMe) strengthened this coordination, while an electron-withdrawing group (p-F) weakened it. Computationally, it was found that weaker coordination by the aromatic group led to a lower activation barrier, as the electrophilicity of the metal is enhanced in the C-C coupling step. Formation of the major enantiomer through this TS was preferentially accelerated, leading to greater predicted enantioselectivity with the fluorinated ligand. A tighter substrate-Rh complexation (a consequence of a slightly weaker ligand-metal interaction) enhances any unfavourable steric effects. This result was validated experimentally across a range of substrates, and was successful in both the matched and mismatched double-stereodifferentiating setting (*i.e.* competing catalyst vs. substrate stereocontrol).

## 5. Challenges in quantifying selectivity

### 5.1 Describing non-covalent Interactions

The growth of non-covalent asymmetric catalysis coincides with an ever-greater appreciation of the challenges associated in accurate computational modelling of non-bonding interactions. In the context of computing stereoselectivity, this may be important in obtaining qualitative geometries and in the evaluation of relative stabilities of competing TS structures. For example, Krenske and Houk have focussed on the importance of aromatic interactions in stereoselective reactions, such as [4 + 3] cycloadditions of oxallyl intermediates with furans reported by Hsung<sup>40</sup>. Here, an Evans chiral auxiliary provides stereoinduction; however, when a phenyl substituted oxazolidinone auxiliary was used the reaction

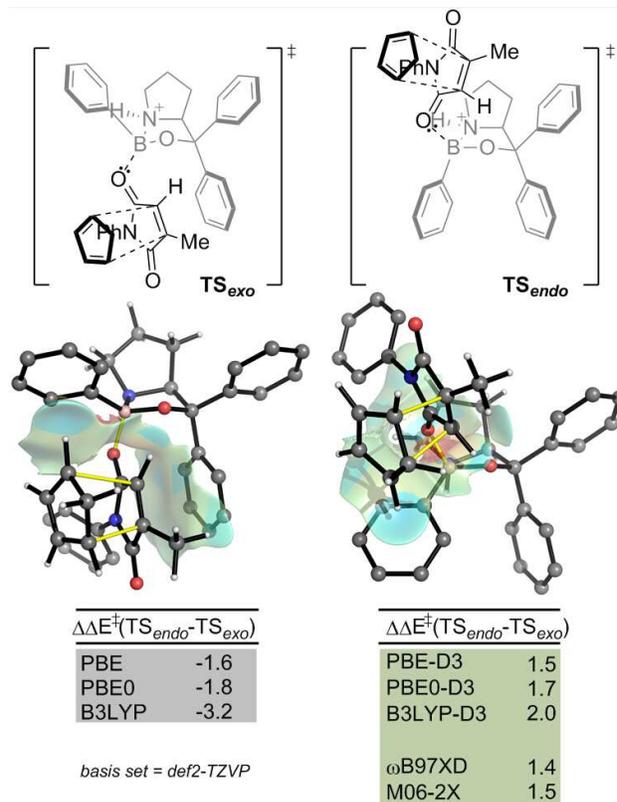
proceeded in a contra-steric sense, the diene approaching the same side as the phenyl group. DFT computations reveal the existence of a stabilising edge-to-face interaction between furan and phenyl group, which stabilises this pathway. Quantitative estimation of the energy difference between the two competing [4 + 3] TSs illustrates the importance of dispersion in this favourable interaction. As can be seen in Fig. 11,  $\Delta\Delta E^\ddagger$  increases from 0.2 kcal·mol<sup>-1</sup> with B3LYP to 1–2 kcal·mol<sup>-1</sup> with dispersion-corrected or M06-2X functionals, since B3LYP effectively neglects this interaction. Uyeda and Jacobsen<sup>41</sup> have discovered a key aromatic interaction between catalyst and substrate in the asymmetric catalysis of Claisen rearrangements. In the most stable TS an aromatic group of the catalyst interacts with a polar C-H group of the substrate, and differing aromatic substitution patterns modulate this interaction and hence enantioselectivity. While B3LYP/6-31G(d) results give poor correlation with the experimental trend, M05-2X/6-31G(d) calculations led to an improvement in the computational performance ( $R^2 = 0.88$  over seven different catalysts) due to a more accurate description of non-covalent interactions.



**Fig. 11.** Relative energies (in kcal·mol<sup>-1</sup>) of TS and TS', computed with different functionals and the 6-31G(d) basis set over the B3LYP/6-31G(d) geometries.<sup>a</sup> In this case the 6-311+G(d,p) basis set was used.

For Cinchona-amine catalysed intramolecular aldol reactions the origins of stereoinduction in the Houk-Lam model<sup>42</sup> lie in the conformational preference of the hydrogen-bonded nine-membered cyclic aldol TS. Multiple non-bonded interactions occur in each competing TS, and the inclusion of dispersion correction to the B3LYP-computed energetics resulted in closer quantitative agreement with experimental selectivities.

For asymmetric catalysis with chiral boron Lewis-acids, such as Corey's oxazaborolidines the importance of so-called "non-classical" hydrogen bonds has been found to be important in pre-organisation of the ensuing TSs<sup>43</sup>. In the case of asymmetric Diels-Alder reactions of maleimides promoted by protonated oxazaborolidine catalysts, we found little computational support for these interactions as stereo-controlling elements as judged by NBO and QTAIM analysis<sup>44</sup>. However, sizable dispersive interactions do occur in the more favourable *exo*-face TS, which leads to the major enantiomer.



**Fig. 12.** Competing *exo*- and *endo*-face coordinated TS structures in an asymmetric Diels-Alder reaction: the inclusion of dispersion is essential to correctly account for the observed enantioselectivity.

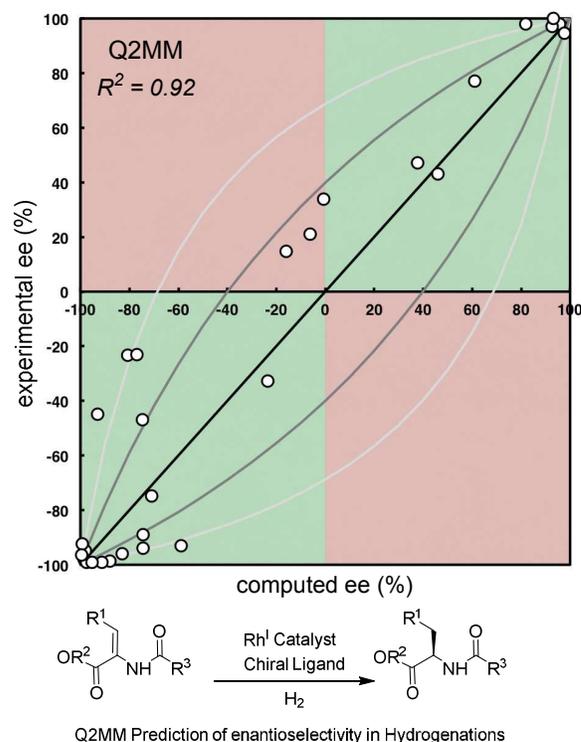
In Fig. 12 these regions are illustrated for the competing TSs by the inclusion of the NCI isosurface, which outlines the region of space where the maleimide sits over the catalyst in the more stable structure. Using DFT-D and M06-2X functionals we achieve a quantitative reproduction of the experimental selectivity (86%, ee  $\Delta\Delta G^\ddagger = 1.1$  kcal·mol<sup>-1</sup> with M06-2X/6-31G(d)). In contrast, the computed sense of selectivity is incorrectly inverted when the same functionals are applied without an explicit dispersion correction.

Medium- to long-range interactions pose a particular problem to DFT due to the semi-local treatment of electron correlation, which underlies all functional forms in use. This makes precisely gauging the significance of non-covalent interactions in stereoselectivity difficult, although as we have described above, scenarios in which DFT-D results give contrasting results implicate an important role. One means to qualitatively assess the extent of non-bonding interactions comes from an NCI analysis as developed by Contreras-Garcia and colleagues<sup>45</sup>. Here, regions of slowly varying electron density corresponding to non-covalent interactions are highlighted graphically: the first example of this technique to an asymmetric reaction was performed by Rzepa and Hii<sup>46</sup>, in the silver-catalysed addition of alcohol and amine nucleophiles to allenes. Using this technique they showed that noncovalent interactions between the chiral ligand and substrate play a key role in determining selectivity.

## 5.2 Conformational Flexibility

The computational study of systems in which a number of conformations are accessible poses a challenge in terms of exhaustively sampling the possible TS geometries. The description of bond making/breaking may be captured by electronic structure theory; however, such calculations are typically incompatible with automated conformational sampling algorithms (e.g. molecular dynamics or Monte Carlo) even at relatively low levels of theory. An active area of research thus surrounds alternative descriptions of TSs using classical approaches, which are amenable to incorporation with standard searching techniques that require multiple optimisations to be performed on-the-fly. In asymmetric reactions and catalysis the most well-developed of these has proven to be quantum-guided molecular mechanics (Q2MM)<sup>9</sup>. The origins of this approach lie in the reaction-specific parameterisation of molecular mechanics force fields to describe a TS geometry, as in Houk's work on radical cyclisations culminating in the Houk-Beckwith model for diastereoselectivity in 5-*exo-trig*-cyclisations<sup>47</sup>. The TS is represented as a minimum by the force field, although key distances, angles and dihedrals are reproduced faithfully. Norrby has pioneered the systematic optimisation of MM parameters based on reference data from QM calculation, and has successfully demonstrated the application of this approach to challenging problems of stereoselectivity prediction in flexible organic and organometallic systems. Studying the rhodium-catalysed hydrogenation of enamides with chiral bisphosphine ligands, the Q2MM approach was tested using a full set of known ligands having varying efficiency in asymmetric hydrogenation, and a range of substrates. An impressive correlation was obtained with only 3 anomalies from 29 calculations, the remainder giving the correct handedness of the product. The correlation between computation and experiment ( $R^2 = 0.92$ ) demonstrates the potential predictive power of such an approach. QM-guided force-field development is under continued development by Norrby and Wiest, such that Q2MM offers the possibility of computationally predicting enantioselectivities on a timescale that would not be possible with QM calculations. The description of a TS as a minimum rather than a saddle-point has limited the accuracy of some force fields, however, we recently reported that true transition state force fields (TTSFF) may be derived in a similar fashion to earlier work to give superior reproduction of the QM energy surface around the TS<sup>48</sup>.

The Q2MM approach enables automated conformational sampling of TS structures, which can be accurate enough to generate a quantitative prediction of selectivity. Nevertheless, alternative approaches seek to combine sampling at the MM level of theory, or even with semi-empirical calculations (as we have done in the case of enantioselective phase-transfer catalysis) with subsequent QM calculations. Here the lower-level conformational sampling is used as a filter through which only the most promising candidate structures (*i.e.* those within a certain energy window) proceed to a more rigorous QM optimisation. Whatever the approach used, the result is an ensemble of TSs leading to the competing products. The selectivity is computed taking into account the contributions of all structures: in Fig. 13, the R-selectivity would be overestimated if only the two most stable TS energies were analysed, since there are two low-lying structures for the S-pathway. The (enantio)-selectivity may be obtained assuming a thermal ensemble involving all TS structures, and by computing the Boltzmann factor for each.



Q2MM Prediction of enantioselectivity in Hydrogenations  
**Fig. 13:** Q2MM computed enantioselectivity against experimental results for asymmetric Rh-catalysed hydrogenation of enamides, from ref<sup>49</sup>. Green areas correspond to correct prediction of the absolute sense of enantioselectivity, whereas red areas correspond to a prediction in the opposite sense to experiment.

## 6. Key Considerations

In the following paragraphs we provide some guidelines to keep in mind when selecting a computational method. There is of course no definitive approach, and the recommended method of choice will usually be a compromise between accuracy and computational cost. Although this review focuses mainly on enantioselective catalysis, the suggestions outlined below can also be applied to other studies.

### 6.1 Choosing the “appropriate” method

The selection of the level of theory will depend on the chemical process under study, *i.e.* the kind of bonds being broken and formed, the influence of nonbonding interactions, and the size and conformational flexibility of the system.

**Optimisation:** For rigid organic molecules, optimised geometries are relatively insensitive (compared with other properties or relative energies) to the choice of basis set or density functional. However, when dispersion effects are important, the level of theory chosen can profoundly affect geometries. This is particularly relevant for flexible molecules, where a range of non-covalent interactions can allow a large number of conformations, some of them very close in energy. In general, the use of a dispersion corrected functional along with a 6-31G(d) basis set is usually enough to capture the geometry of the species of interest.

**Energetics and other electronic properties:** To obtain reliable estimates on these quantities higher-level calculations are

recommended. This can be done, for example, using a triple- $\zeta$  quality basis set on the previously optimised structures. Quantitative estimates of selectivity rely upon relative energy differences, militating against systematic errors in the computation of barriers of energy changes. For stereoisomers there does appear to be beneficial error cancellation such that accuracy of less than 1 kcal·mol<sup>-1</sup> can commonly be achieved, as is the case for the examples mentioned above. This might explain in part the success of the B3LYP functional on the study of several asymmetric processes. However, it is difficult to know this in advance, and other DFT functionals or WF approaches may be more appropriate in some cases. Guidance for choosing a DFT functional can be found from previous benchmarking studies for related chemical systems. Comparisons to gas phase CCSD(T)/CBS may be possible for small systems. However, comparison with experimental selectivity values is probably more useful.

Finally, while thermochemical properties including enthalpies, entropies, and heat capacities may be important in order to compare to experimental data, the accurate calculation of these quantities still represents a severe challenge. These values can usually be obtained within the harmonic approximation; however they are known to be extremely sensitive to the geometries, basis set, optimisation convergence criteria and grid. Moreover, they can be notoriously inaccurate when several low-lying vibrational frequencies *i.e.* less than 100 cm<sup>-1</sup> exist. In such cases, relative free energies may be less reliable than relative energies. Several approaches have been included to estimate entropies, including the treatment of low-lying vibrational modes by a free-rotor approximation, and empirical corrections to account for solvent effects.

## 6.2 Solvent effects

In many cases, the effect of solvent can be considered via a simple implicit solvent model calculation (*e.g.*, PCM, CPCM, SMD) coupled with UFF, UAHF, or UAKS radii to define the solute cavity. While solvation can be included as a correction on the energy of the system, one should also consider its inclusion in optimisation calculations, particularly when charged and/or highly polarizable species are involved. While the use of implicit solvent models has proven adequate in many applications, the use of explicit solvent may be relevant in cases when specific interaction between the solvent and substrate are important. Computational studies on stereoselectivity in explicit solvent are still very rare. However, with the development of new techniques and the increase in computational power, they are expected to become more accessible.

## 6.3 Configurational Space

Current approaches provide a clear protocol to find local energy minima on the PES. For small organic molecules it may be relatively easy to manually inspect the accessible configurations. However, for larger systems a more efficient (preferably automated) conformational search is required in order to identify the conformations accessible under the experimental conditions. Following from this, for each conformation the energy has to be calculated and Boltzmann-weighted to obtain the predicted energy and or selectivity ratio. For asymmetric catalysis, this includes exploring the conformations of substrates and catalyst, as well as the different binding modes between them, which also determine

the outcome of the reaction. MM-based conformational searches may be computationally affordable but their reliability is dependent on the force field parameters used. In the absence of specific parameterization and testing for a system of interest, the quality of structures and energetics is highly uncertain. The MM configurational sampling of transition states is possible via the Q2MM method. Recent advances in the automated following of reaction coordinates with QM calculations suggest that this will become more prominent in future years.

## 6.4 Validation

In the comparison of theoretical results against experiment it is helpful to appreciate the inherent accuracy of the chosen computational methodology – *i.e.*  $\pm 1$  kcal·mol<sup>-1</sup> is not uncommon for DFT-computed relative energies, while barrier heights and energy changes will be less accurate. In calculations of stereoselectivity, good quantitative agreement, particularly for only a single data point, may be coincidental. Reliance on multiple experiments – including negative data and repeated measurements – is helpful to ensure that computational results are “right for the right reason”. Furthermore, we recommend computational work should be performed blind *i.e.* without prior knowledge of the experimental selectivity to avoid a positive confirmation bias. The ultimate test of the computations is the ability to predict results prior to experiment!

## Conclusions and perspectives

Based on a thorough mechanistic understanding it is possible to quantify levels of stereoselectivity through the use of relatively affordable computational methods. The quantitative results of these calculations may be used to develop understanding and to formulate models, which account for experimental observations. They may also help to predict the outcomes of future experiments: these predictions range from the qualitative to the quantitative, where computation can assist in the design of new asymmetric catalysts. We hope that this review will aid and encourage the organic chemistry community in adopting computations in the study and design of asymmetric reactions.

## Acknowledgements

This research was supported financially by the Royal Society (RG110617 to RSP and a Newton Fellowship to F.D.), the European Community (FP7-PEOPLE-2012-IIF under grant agreement 330364 and 912364 to QP), and CSA-trust grant to QP. We acknowledge the use of the EPSRC UK National Service for Computational Chemistry Software (NSCCS) at Imperial College London and contributions from its staff in carrying out this work (CHEM773). RSP acknowledges the contributions of group members and experimental collaborators past and present, whose work is cited in the references.

## References

1. D. Balcells and F. Maseras, *New J. Chem.*, 2007, **31**, 333-343.

## ARTICLE

## Journal Name

2. G. J. Cheng, X. H. Zhang, L. W. Chung, L. P. Xu and Y. D. Wu, *J. Am. Chem. Soc.*, 2015, **137**, 1706-1725.
3. R. B. Woodward and H. Baer, *J. Am. Chem. Soc.*, 1944, **66**, 645-649.
4. D. G. Truhlar, B. C. Garrett and S. J. Klippenstein, *J. Phys. Chem.*, 1996, **100**, 12771-12800.
5. S. R. Hare and D. J. Tantillo, *Beilstein J. Org. Chem.*, 2016, **12**, 377-390.
6. B. K. Carpenter, J. N. Harvey and A. J. Orr-Ewing, *J. Am. Chem. Soc.*, 2016, **138**, 4695-4705.
7. R. E. Gawley, *J. Org. Chem.*, 2006, **71**, 2411-2416.
8. J. M. Brown, *Organometallics*, 2014, **33**, 5912-5923.
9. E. Hansen, A. R. Rosales, B. Tutkowsky, P.-O. Norrby and O. Wiest, *Acc. Chem. Res.*, 2016, **49**, 996-1005.
10. Y.-h. Lam, M. N. Grayson, M. C. Holland, A. Simon and K. N. Houk, *Acc. Chem. Res.*, 2016, **49**, 750-762.
11. N. T. Anh and O. Eisenstein, *Nouv. J. Chim.*, 1977, **1**, 61-70.
12. L. E. Rush, P. G. Pringle and J. N. Harvey, *Angew. Chem. Int. Ed.*, 2014, **53**, 8672-8676.
13. J. P. Perdew, A. Ruzsinszky, L. A. Constantin, J. W. Sun and G. I. Csonka, *J. Chem. Theory Comput.*, 2009, **5**, 902-908.
14. L. Simón and J. M. Goodman, *Org. Biomol. Chem.*, 2011, **9**, 689-700.
15. S. T. Schneebeli, M. L. Hall, R. Breslow and R. Friesner, *J. Am. Chem. Soc.*, 2009, **131**, 3965-3973.
16. S. Schenker, C. Schneider, S. B. Tsogoeva and T. Clark, *J. Chem. Theory Comput.*, 2011, **7**, 3586-3595.
17. A. Armstrong, R. A. Boto, P. Dingwall, J. Contreras-Garcia, M. J. Harvey, N. J. Mason and H. S. Rzepa, *Chem. Sci.*, 2014, **5**, 2057-2071.
18. H. Kruse, L. Goerigk and S. Grimme, *J. Org. Chem.*, 2012, **77**, 10824-10834.
19. S. Bahmanyar, K. N. Houk, H. J. Martin and B. List, *J. Am. Chem. Soc.*, 2003, **125**, 2475-2479.
20. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215-241.
21. G. A. DiLabio and M. Koleini, *J. Chem. Phys.*, 2014, **140**, 18A542.
22. S. Grimme, *WIREs Comput. Mol. Sci.*, 2011, **1**, 211-228.
23. L. X. Yang, C. Adam, G. S. Nichol and S. L. Cockroft, *Nat. Chem.*, 2013, **5**, 1006-1010.
24. B. List, R. A. Lerner and C. F. Barbas, *J. Am. Chem. Soc.*, 2000, **122**, 2395-2396.
25. D. W. C. MacMillan, *Nature*, 2008, **455**, 304-308.
26. D. Parmar, E. Sugiono, S. Raja and M. Rueping, *Chem. Rev.*, 2014, **114**, 9047-9153.
27. C. Allemann, J. M. Um and K. N. Houk, *J. Mol. Catal. A: Chem.*, 2010, **324**, 31-38.
28. R. R. Knowles and E. N. Jacobsen, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 20678-20685.
29. T. Akiyama, J. Itoh, K. Yokota and K. Fuchibe, *Angew. Chem. Int. Ed.*, 2004, **43**, 1566-1568.
30. D. Uraguchi and M. Terada, *J. Am. Chem. Soc.*, 2004, **126**, 5356-5357.
31. L. Simón and J. M. Goodman, *J. Am. Chem. Soc.*, 2008, **130**, 8741-8747.
32. J. P. Reid, L. Simón and J. M. Goodman, *Acc. Chem. Res.*, 2016, **49**, 1029-1041.
33. T. Marcelli, P. Hammar and F. Himo, *Chem. Eur. J.*, 2008, **14**, 8562-8571.
34. R. S. Paton and J. M. Goodman, *J. Chem. Inf. Model.*, 2009, **49**, 944-955.
35. I. D. Gridnev, M. Kouchi, K. Sorimachi and M. Terada, *Tetrahedron Lett.*, 2007, **48**, 497-500.
36. K. N. Houk and P. H. Y. Cheong, *Nature*, 2008, **455**, 309-313.
37. A. D. G. Yamagata, S. Datta, K. E. Jackson, L. Stegbauer, R. S. Paton and D. J. Dixon, *Angew. Chem. Int. Ed.*, 2015, **54**, 4899-4903.
38. R. N. Straker, Q. Peng, A. Mekareeya, R. S. Paton and E. A. Anderson, *Nat. Commun.*, 2016, **7**, 10109.
39. M. C. Nielsen, K. J. Bonney and F. Schoenebeck, *Angew. Chem. Int. Ed.*, 2014, **53**, 5903-5906.
40. E. H. Krenske and K. N. Houk, *Acc. Chem. Res.*, 2013, **46**, 979-989.
41. C. Uyeda and E. N. Jacobsen, *J. Am. Chem. Soc.*, 2011, **133**, 5062-5075.
42. Y.-h. Lam and K. N. Houk, *J. Am. Chem. Soc.*, 2015, **137**, 2116-2127.
43. R. C. Johnston and P. H.-Y. Cheong, *Org. Biomol. Chem.*, 2013, **11**, 5057-5064.
44. R. S. Paton, *Org. Biomol. Chem.*, 2014, **12**, 1717-1720.
45. J. Contreras-Garcia, E. R. Johnson, S. Keinan, R. Chaudret, J. P. Piquemal, D. N. Beratan and W. T. Yang, *J. Chem. Theory Comput.*, 2011, **7**, 625-632.
46. J. L. Arbour, H. S. Rzepa, J. Contreras-García, L. A. Adrio, E. M. Barreiro and K. K. Hii, *Chem. Eur. J.*, 2012, **18**, 11317-11324.
47. J. E. Eksterowicz and K. N. Houk, *Chem. Rev.*, 1993, **93**, 2439-2461.
48. Á. Madarász, D. Berta and R. S. Paton, *J. Chem. Theory Comput.*, 2016, **12**, 1833-1844.
49. P. J. Donoghue, P. Helquist, P. O. Norrby and O. Wiest, *J. Chem. Theory Comput.*, 2008, **4**, 1313-1323.