Journal of Molecular Catalysis A: Chemical 324 (2010) 146–155

Contents lists available at ScienceDirect

Journal of Molecular Catalysis A: Chemical

journal homepage: www.elsevier.com/locate/molcata

Computational catalysis

Theory and application of medium to high throughput prediction method techniques for asymmetric catalyst design☆

Christopher R. Corbeil[a,b], Nicolas Moitessier[a,*]

[a] Department of Chemistry, McGill University, 801 Sherbrooke Street W, Montréal, Québec, Canada H3A 2K6
[b] Biotechnology Research Institute, National Research Council Canada, 6100 Royalmount Avenue, Montreal, Quebec, Canada H4P 2R2

ARTICLE INFO

Article history:
Received 4 January 2010
Accepted 16 March 2010

Keywords:
Asymmetric catalysts
HTVS
QM/MM
Q2MM
QSSR
Molecular mechanics
Virtual screening
ACE

ABSTRACT

With the use of computational methods in the field of drug design becoming ever more prevalent, there is pressure to port these technologies to other fields. One of the fields ripe for application of computational drug design techniques; specifically virtual screening and computer-aided molecular design, is the design and synthesis of asymmetric catalysts. Such methods could either guide the selection of the optimal catalyst(s) for a given reaction and a given substrate or provide an enriched selection of highly efficient asymmetric catalysts which enable the synthetic chemists to focus on the most promising candidates. This would in turn provide savings in time and reduce the costs associated with the synthesis and evaluation of large libraries of molecules. However, to be applicable to the evaluation of a large number of potential catalysts, speed is of utmost importance. This impetus has led to the development of medium to high throughput virtual screening (HTVS) methods for asymmetric catalyst development or assessment, although a very few applications have been reported. These methods typically fall into four classes: methods combining quantum mechanics and molecular mechanics (QM/MM), pure molecular mechanics-based methods – a class which can be subdivided into static and dynamic transition state modeling – and lastly quantitative structure selectivity relationship methods (QSSR). This review will cover specific methods within these classes and their application to selected reactions.

Contents

1. Introduction

Financial and environmental pressures in the drug discovery and development field require that novel drugs are found in a time and cost-effective manner. To fulfill these requirements, computational techniques have found their way into the toolkit of medicinal chemists providing a viable alternative and/or complement to experimental approaches such as high throughput screening [1–3]. In fact, there are now many fairly predictive methods (e.g., QSAR, docking programs, combinatorial library profiling) available to drug discovery and development chemists [4]. Although these methods are based on approximations, they are accurate enough to yield a higher rate of finding lead molecules when screening large libraries compared to the traditional experimental approaches [5–7]. Even though these techniques provide small libraries enriched in bioactive molecules, the small number of potentially missed bioactive molecules, does not often outweigh the speed and cost savings of screening a library *in silico*.

☆ This paper is part of a special issue on computational catalysis.

* Corresponding author. Tel.: +1 514 398 8543; fax: +1 514 398 3797.
E-mail address: nicolas.moitessier@mcgill.ca (N. Moitessier).

1381-1169/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.molcata.2010.03.022

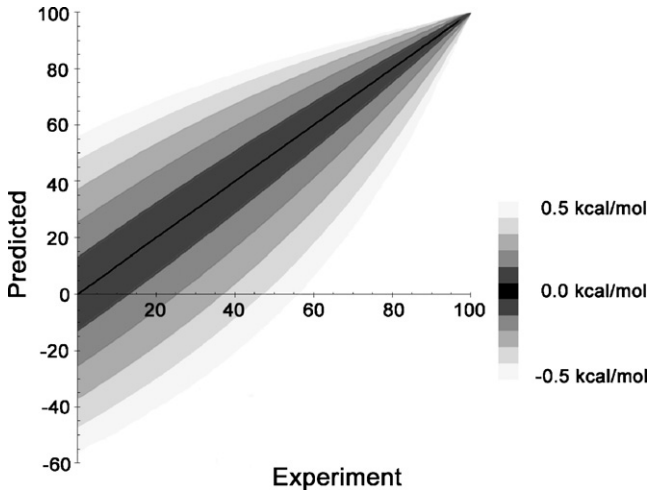

Fig. 1. Predicted versus experimental stereomeric excess. Gradient shows error in % stereomeric excess in relation to error associated with the prediction of TS energy at −78 °C.

Unfortunately, these advances and successes in the field of computational drug design and development have not stimulated the development of virtual screening (VS) or computer-aided design methods in other chemistry disciplines such as in the design of asymmetric catalysts. Computational tools in the field of asymmetric catalyst design are typically exploited to rationalize the outcome of a given reaction *post facto* rather than to predict its outcome *a priori*. Yet, the ability to predict the stereomeric excess of a reaction would enable organic chemists to quickly test out new asymmetric catalyst structures and to prioritize a few of them for synthesis or applications to specific substrates.

The lack of quick predictive computational tools for organic chemists when compared to the field of drug design and development is attributed to one major contributing factor: these tools require a higher accuracy than those found in the field of drug design. To accurately discriminate an excellent from a poor asymmetric catalyst, it is necessary to have an error for the predicted transition state (TS) free energy differences (necessary to compute the stereomeric excess), in the range of 0.5 kcal/mol or less (see Fig. 1). On the other hand, discriminating drug hits from non-binders requires a lower resolution in the order of 3–5 kcal/mol and is calculated on ground state structures.

Virtual drug screening methods such as docking programs use scoring functions to predict the ligand binding affinity, with many methods using force fields to calculate the potential energy of the ligand–protein complex [4,8]. However, molecular mechanics force fields have been developed to simulate the ground states of molecules and cannot be applied directly to the computation of TS energies. To determine TS structures and associated energies, the most accurate approach remains quantum mechanics (QM). However, although QM can compute TS structures and energies very accurately [9], it still lacks the speed required for the development of a QM-based VS tool and can hardly be applied to large catalytic systems [10–17]. In addition, using QM requires expert knowledge for the selection of the correct basis set. This necessary expertise together with the required CPU resources is a major hurdle for the use of QM-guided catalyst design by experimentalists. To address these issues, computational organic chemists have developed methods and programs that enable a faster calculation of stereomeric excesses or prediction of favoured stereoisomers [18–20].

Methods for the prediction of stereomeric excess typically differ in one or all of the following four areas: (1) consideration of the complex (catalyst/substrate) flexibility, (2) consideration of the reacting center flexibility, (3) computation of the complex (catalyst/substrate) interactions or energy and (4) computation of the energy of the reacting center. Even with the above variables the reported methods typically fall into four classes: methods combining quantum mechanics and molecular mechanics (QM/MM), pure molecular mechanics-based methods (which can be subdivided into rigid and dynamic TS models) and quantitative structure selectivity relationship (QSSR). Herein, we will describe a selection of methods that have been applied to organic asymmetric catalysts. Each of these methods, their pros and cons, will be illustrated by selected examples. This review is by no means an exhaustive review but rather a discussion on the existing methods. We will focus on how they treat the flexibility and energy of both the reacting center and/or the catalyst/substrate complex as a whole. Although some of the methods mentioned herein have not been directly applied to organic asymmetric catalysts, they can yet be easily transferable to this field and will be mentioned.

2. Virtual screening of asymmetric catalysts

2.1. Using quantum methods for virtual screening of asymmetric catalysts

Quantum mechanics has been exploited to rationalize experimental results and provide valuable insight into the reaction pathway of many reactions [18,19,21]. Great care must be taken when selecting basis sets for QM methods since smaller set (i.e., quicker methods) may provide qualitative answers but not the desired quantitative predictions. It has only been recently that a single study undertook a screening of a large series of catalytic systems. Schneebeli et al. [11] computationally screened 46 dioxirane catalyzed epoxidation reactions (Fig. 2) and demonstrated good correlation between predicted and experimental stereomeric excesses with a mean unsigned error (MUE) of around 0.5 kcal/mol and 20%ee. The MUE is a measure of the average absolute error between the experimentally determined and predicted values. They also demonstrated the sensitivity of the results to the selection of the basis set. However, although QM can calculate the TS structures and energies very accurately [10–17], the need for expertise in both the method used and the reaction under investigation when using QM-based methods limit their applicability for synthetic chemists. In addition, this method can hardly be applied to flexible systems. In fact, the epoxidation catalysts studied were all cyclic (i.e., rigid) catalysts.

To address the computational time issue, it is possible to use a hybrid technique, called QM/MM [22], which treats the atoms involved in the bond breaking and forming with QM and the rest of the molecule with molecular mechanics (MM). In the context of prediction of stereomeric excess and TS structures, QM/MM hybrid approaches are applicable to large systems and are relatively quicker than QM methods alone while still modeling *ab initio* the

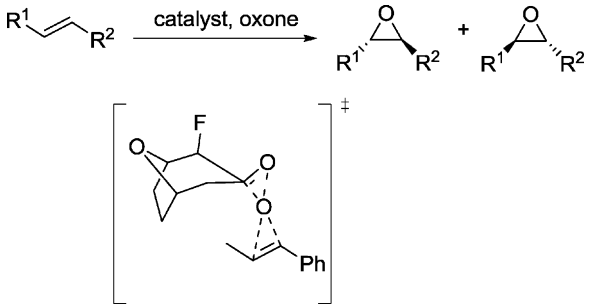

Fig. 2. Dioxirane catalyzed epoxidation reaction mechanism. General synthetic scheme (top) and transition state for a selected example (bottom).

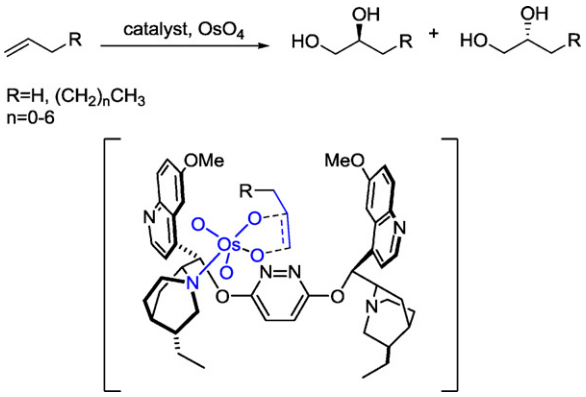

Fig. 3. QM/MM study of the asymmetric dihydroxylation of alkenes (QM in Blue, MM in black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

core of the TS [23]. However, a conformational search of the whole system should be carried out separately prior to the computation of the derived optimized conformations.

One of the only examples of QM/MM methods applied to the prediction of stereoselectivity is that of the dihydroxylation of alkenes (Fig. 3) [24]. Initial studies using styrene as a substrate and experimentally determined catalyst structures demonstrated the predictive power of QM/MM applied to this reaction, with a predicted stereoselectivity closely matching the experimental results (99.4%ee predicted, 96%ee observed) [25]. In this first study, they assume that the solution structures of the catalyst and substrate remain unchanged when in complex.

With this promising result, Ujaque et al. then considered more flexible substrates. However, studying *n*-alkenes using traditional QM methods would prove to be too difficult due to the explosion of possible conformations when going from propene to 1-decene. To address this combinatorial explosion, the authors carried out an initial systematic search of all possible conformations of the *n*-alkenes in complex with a rigid catalyst. In this study only two conformations of the catalysts that have been previously found to be favoured, were considered [25]. This conformational search was carried out in two steps. First, 12 conformations of the OsO_4•NH_3 + CH_2=$CHCH_3$ TS core were initially generated accounting for the three possible approaches of the olefin and four possible orientations of the olefin upon approach (two leading to the (*R*) and two leading to the (*S*) enantiomer). They then proceeded with a systematic conformational search of the alkene/catalyst systems separately on each conformation of the core which was frozen. From this set of conformations, up to 300 of the lowest energy conformations for each catalyst–substrate complex according to MM potential energies were selected. Finally, each of these conformations, including the TS core, underwent full optimization using the selected QM/MM method and stereoselectivities were derived. The stereoselectivities were predicted with reasonable correlation with experimental results, agreeing with the increase in stereoselectivity following the increase in chain length. The stereoselectivities were predicted with a MUE of about 0.25 kcal/mol and under 10%ee. In this study, it clearly appears that the protocol cannot be easily applied by non-expert once more limiting the applicability of this technique by synthetic chemist on a daily basis.

2.2. Rigid transition state molecular mechanics models

QM methods have shown promises as tools to design novel asymmetric catalysts. However, QM or even QM/MM is still significantly too slow to be advantageously used to screen or design novel structures as compared to experimental stepwise optimization or screening. To overcome this issue one could envision the use of MM exclusively [20,21]. MM offers the advantage of being significantly quicker although with some caveats. Traditional MM force fields have been developed to predict the ground state structures and energies of molecules and cannot consider breaking or forming bonds. It is therefore necessary to derive FF parameters for TSs. The second issue is the conformational search of the catalyst/substrate systems. While conformational analysis is usually considered separately when using QM or QM/MM methods, conformational search engines are directly implemented in MM programs reducing the necessary steps of the MM-based protocols. However, these engines locate the global (or near global) minima of the energy function which, when using available force fields, correspond to the ground state of a molecule. As a result, these engines will never identify a TS located on a saddle point of the energy function.

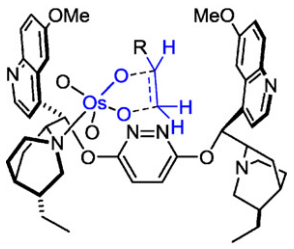

Fig. 4. Example of rigid TS mode. In blue are the reacting atoms which would be frozen during optimization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

To address these two major issues, transition state force fields (TSFF) model TSs of a reaction as a minimum on a potential energy surface (PES). The simplest TSFF freezes or constrains the breaking or forming bonds and their associated angles and dihedrals (Fig. 4). A model system is first developed using QM or crystallographic methods [21]. This model is then used to derive the optimum values for bond lengths, angles and dihedrals. These interactions can then be added to the force field with very large force constants. Any movement away from the optimum values are assigned high penalties hence constraining all atoms to the transition state geometry. With this MM approach, conformational searches can be carried out and the TS energies can be computed. This model is sound as no significant change in geometry of the transition state is usually observed from one catalyst and/or reactant to the next while the rest of the system which is not involved in the reaction can be assumed in its ground state.

A theoretical study of hydroborations by Houk et al. (Fig. 5) [26] was the first application of a TSFF. A model system of the reacting center consisting of ethylene and BH_3 was computed using HF/3-21G*. From this model were derived MM2 parame-

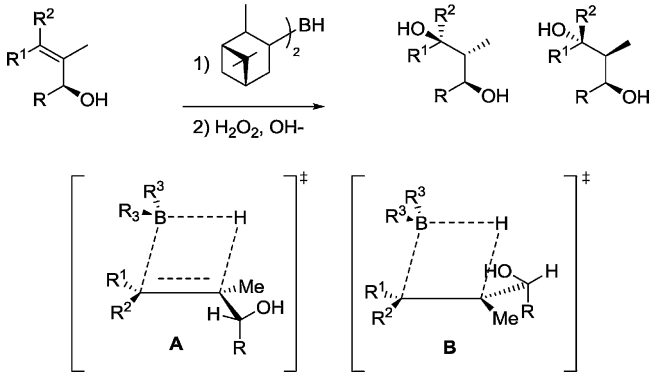

Fig. 5. Hydroboration of alkenes.

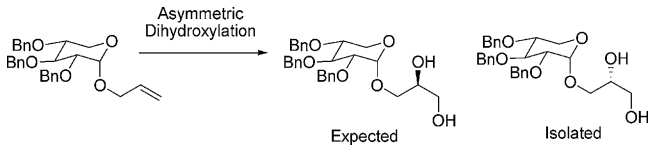

Fig. 6. Sharpless asymmetric dihydroxylation of xylose.

ters subsequently used to constrain the atoms involved in the TS. Chiral boranes reported in the literature were next built and a conformation search using the modified MM2 (encompassing the constrained TS parameters) was applied. If multiple conformations with similar energies were found, a Boltzmann distribution over all conformations was used to compute the stereoselectivities. This pioneering study showed that MM can be applied to predict the stereoselectivity of a reaction with good accuracy, with an MUE of around 0.5 kcal/mol and 30%ee. Even though this approach is expected to be less accurate than QM methods, the ability to screen compounds with higher throughput made it applicable to the VS of new catalysts.

Moitessier et al. also used a rigid TS model TSFF to aid in the rationalization of the unexpected outcome of the dihydroxylation of a benzyl protected allyl xylopyranoside. In this study, the isolated major isomer was opposite to the one expected from the Sharpless mnemonic (Fig. 6) [27,28].

An initial model from a previous QM study [29] was exploited to build the reacting center core, while using a modified CFF91 force field for the optimization of the catalyst/substrate complexes. To account for the flexibility of the catalyst/substrate complex a genetic algorithm was used [8]. When applied to the dihydroxylation of the benzyl protected xylopyrannoside (**1**, Fig. 7), it showed that unexpected isomer was formed due to alkene being too large to adopt either binding mode described previously by the Corey and Sharpless model (**2** and **3**, Fig. 7). With these promising results, a VS of alkenes was undertaken. Although the accuracy of the predicted stereoselectivities was lower than pure QM methods (MUE of around 0.7 kcal/mol and 15%ee), the method competed within a fraction of the time. In fact, this method was able to distinguish the substrates providing the best stereoselectivities showing its promise as a method for VS.

Another technique has been developed by Harriman and Deslongchamps and termed reverse docking [30]. While traditional docking is where a ligand is docked flexibly into a rigid protein (Fig. 8A), a rigid transition state can be docked into a flexible catalyst (Fig. 8B), modeling the reactant/catalyst transition structure. The resulting structures and associated energies can then be used to predict stereoselectivities.

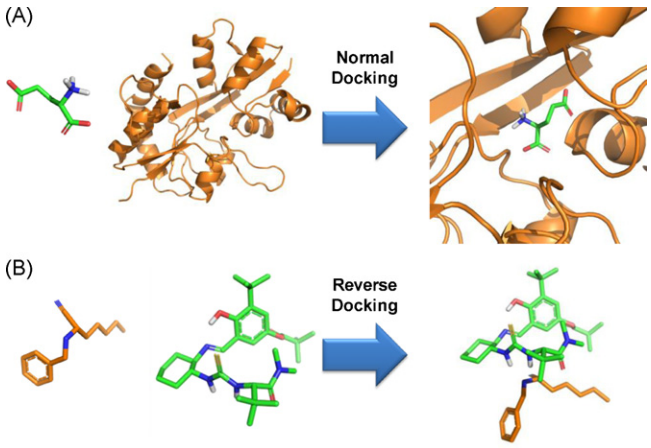

Fig. 8. Normal versus reverse docking methods for reaction shown in Fig. 9A. (A) In normal docking a flexible ligand (green) is docked into a rigid protein (orange). (B) In reverse docking the rigid TS is docked into the flexible catalyst (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

The initial reverse docking method was built around the AutoDock [31] conformational search engine and energy function along with a TS optimized using HF/6-31G*. The substrate initially undergoes a conformational search using the force field MMFF94s. The lowest energy conformation for the substrate is then used as input for a QM calculation to derive the TS. The TS substrate is then frozen and the catalyst is added to the system and undergoes a conformational search using AutoDock. This method is therefore limited to reactions in which the catalyst is not involved as part of the TS. For instance, in the dihydroxylation shown above, the reacting center includes the substrate (i.e., styrene), catalyst (i.e. $(DHQD)_2PHAL$) and reagent (OsO_4) and could not be treated using reverse docking.

Application of reverse docking to the azidation of α,β-unsaturated carbonyls with Miller's catalyst validated the approach (Fig. 9A). This first version was able to predict the conformation of the catalyst and the favoured stereoisomer although the stereomeric excesses were poorly reproduced with MUEs above 5 kcal/mol. The developers then modified the approach to replace

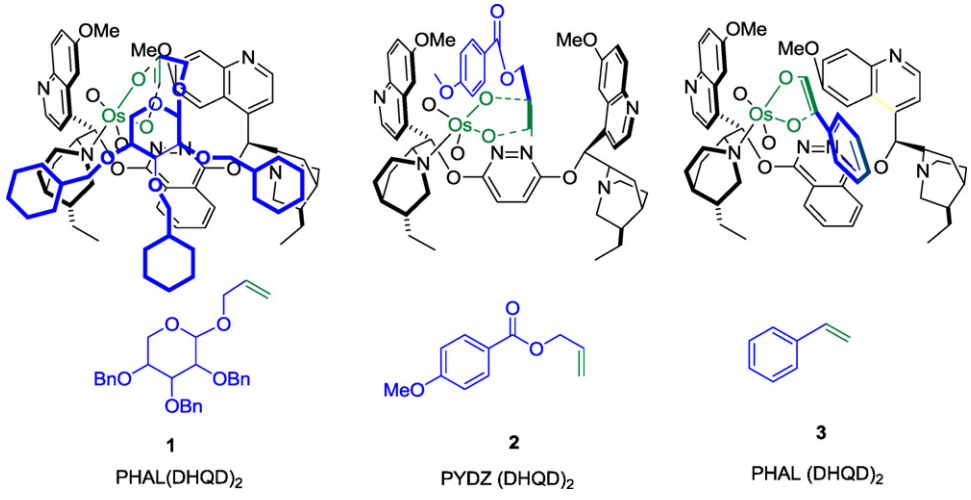

Fig. 7. Sharpless dihydroxylation studied for optimization and validation of genetic algorithm (black, catalyst; blue, alkene; green, frozen atoms). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

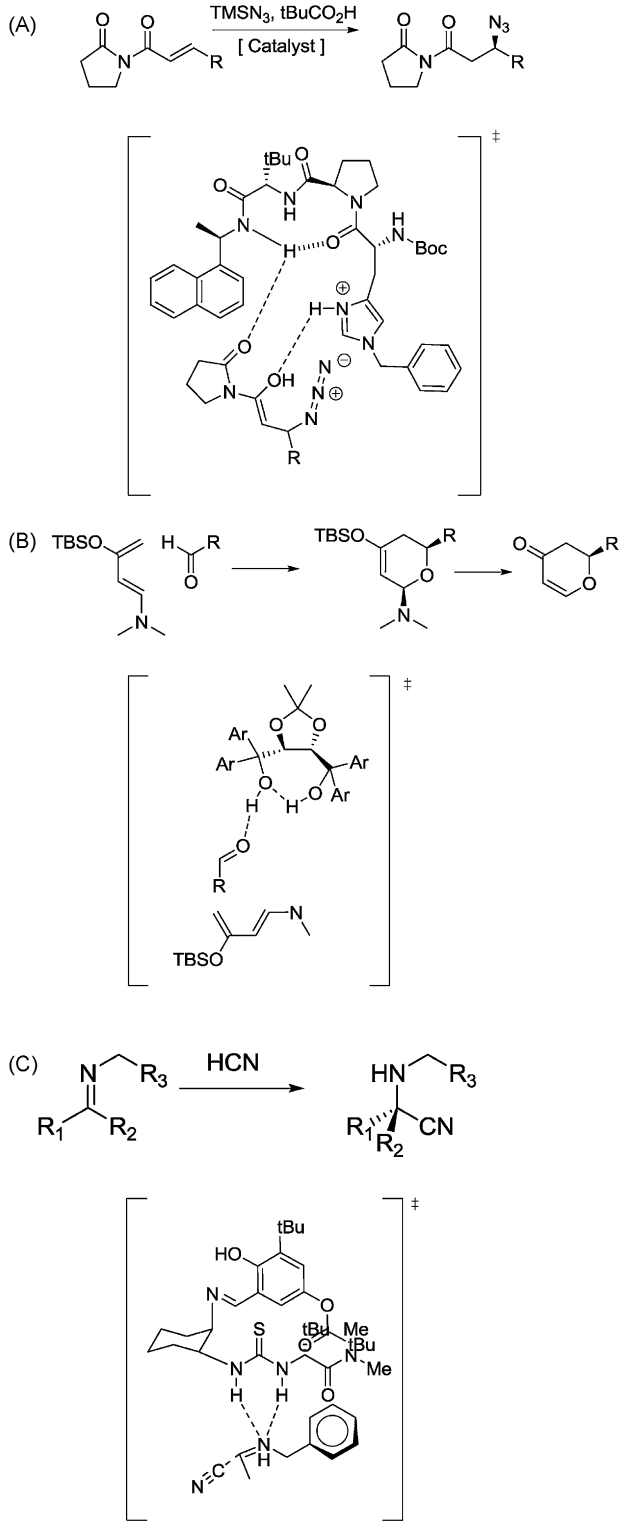

Fig. 9. Reaction studies with reverse docking: (A) azidation of α,β-unstaturated carbonyls, (B) TADDOL-catalyzed asymmetric hetero Diels–Alder reaction and (C) organocatalyzed asymmetric Strecker hydrocyanation of aldimines.

AutoDock with an independent search algorithm program called EM-Dock, which has recently been implemented within MOE [32]. Like AutoDock, EM-Dock calls for a genetic algorithm to perform the conformational search. The main difference between the AutoDock and EM-Dock approaches is their representation of the energy function for intermolecular interactions. While Autodock pre-computes the interaction energy on grids using various probes, EM-Dock uses a pair-wise potential to calculate these intermolecular interactions. This second version, applied to the TADDOL-catalyzed asymmetric hetero Diels–Alder reaction (Fig. 9B), identified the favoured stereoisomer yet did not yield an improvement in the absolute prediction of stereoselectivity [32]. After some more modifications of the algorithm, Deslongchamps and co-workers revisited the TADDOL-catalyzed asymmetric hetero Diels–Alder reaction [33] and also applied EM-Dock to the organocatalyzed asymmetric Strecker hydrocyanation of aldimines and ketimines [34] (Fig. 9C). This enhanced version resulted in the desired increase in accuracy. EM-Dock yielded predictions of stereoselectivity with a MUE of around 0.7 kcal/mol and around 5%ee for the TADDOL-catalyzed asymmetric hetero Diels–Alder reaction for catalysts with higher than 90%ee [33]. When examining the organocatalyzed asymmetric Strecker hydrocyanation of aldimines and ketimines [34], EM-Dock was able to identify the correct enantiomer with MUEs of around 1.0 kcal/mol and 35%ee.

2.3. Dynamic transition state molecular mechanics models

In all these MM studies, the atoms involved directly in the formation of new interactions were frozen or constrained, an approximation which is often reasonable. However, when the catalyst conformation changes drastically in the presence of a given substrate or when the catalyst structures are very different, the geometry of the reacting center may move away from the geometry of the model system used to derive the core. For optimal predictions, dynamic TS models allow the geometry of the TS core to freely move while optimizating the geometry. This approach is closer to the search for saddle points with QM. In contrast to the rigid TS model, both conformational search and calculation of the TS core energy using MM is to be carried out simultaneously with the rest of the catalytic system.

An extension of the rigid TSFF would consider smaller force constants for interactions which would allow for some movement away from the optimum equilibrium TS values. The challenge is therefore to derive the proper values of these force constants. MMX was developed so that the equilibrium bond length and force constants are a function of bond order but has never been applied to asymmetric reactions [35]. An issue arises since bond orders are not explicitly known for transition states and in reality force constants may not be directly proportional to bond order. ReaxFF is a similar method which allows the bond order to vary as a function of the bond distance [36–38]. Typically the ReaxFF force field undergoes rigorous training using QM, followed by using molecular dynamics to study a reactive system. Although this technique has mainly been applied to transition metal catalytic system, it has not been applied to asymmetric reactions, yet one could easily envision retraining of the ReaxFF force field for this specific purpose.

To overcome the problem of the definition of bond order mentioned in the above methods, Norrby [39] developed the Q2MM method where the TSFF is entirely developed from QM calculations. This method has been applied to many reactions and has shown to be highly accurate for the prediction of stereomeric excesses [39–49]. Although the use of QM to derive parameters for each new reaction requires a significant investment in time and expertise, once the force field is developed, it allows for quick computations of large libraries of catalysts and/or substrates. Since Q2MM is essentially deriving a force field, any conformational search algorithm can be used to address the flexibility of the system. As an example, Q2MM has been applied to the asymmetric dihydroxylation reaction (Fig. 6) for the prediction and rationalization of stereoselectivities [43–46]. Once the force field has been parameterized for the asymmetric dihydroxylation reaction, a conformational search was performed with each catalyst/substrate system. The search procedure consisted of an initial Monte Carlo

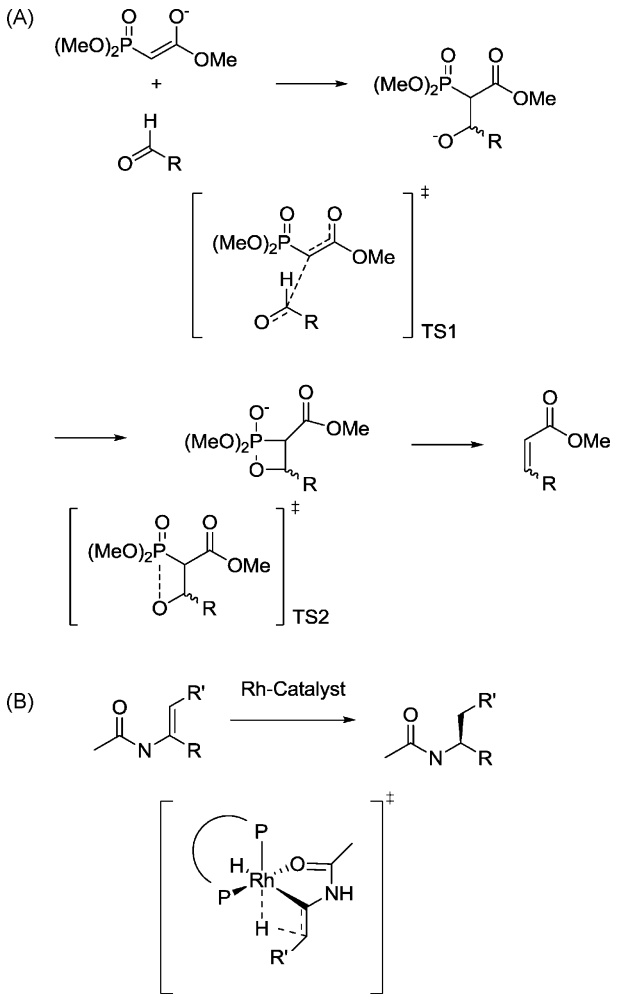

Fig. 10. Examples of reactions studied with Q2MM.

search of "important" rotatable bonds, followed by a more rigorous Monte Carlo/Low Mode search using all torsions. By using Q2MM, good correlations between predicted and experimental selectivities were achieved with an MUE of around 0.7 kcal/mol and under 10%ee [46]. Q2MM has also been successfully applied to the Horner–Wadsworth–Emmons reaction [39,47] (Fig. 10a). This reaction involved two transition states, necessitating the development of parameters for both transition states and the study of multiple diastereomeric pathways to identify the rate-limiting step. In this specific case, *E*/*Z* ratios and not stereomeric excess were predicted. Based on the inability to accurately determine the energy difference between TS1 and TS2, predictions of selectivities were challenging allowing only accurate predictions for catalysts with high selectivity for either *E* or *Z* conformations (i.e., above 95:5) [47]. Q2MM has also been applied to rhodium catalyzed hydrogenation of enamides [48,49]. Previous computational studies have suggested that the addition of the first hydride forming an alkyl hydride complex (see Fig. 10B) was irreversible. The Q2MM investigation therefore focused on this step and its corresponding TS [50–52,10]. Overall, on three Rh-catalysts and a variety of enamides, Q2MM yielded MUEs of approximately 0.65 kcal/mol and around 12%ee.

Another option is to approximate the TS as the intersection of two or more ground states interacting through a mixing term. This mixing term describes the mixing between the reaction and product PES. This technique, known as the empirical valence bond method (EVB) (Fig. 11), was introduced and developed by Warshel

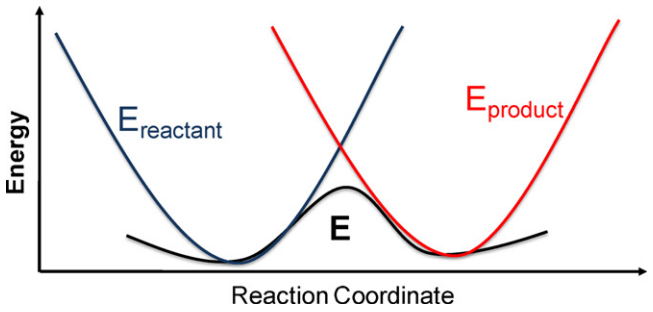

Fig. 11. Mixing of two ground states to find the TS.

and Weiss [53–55]. For the sake of illustration, we will consider only two states (reactants and products), but more than two states can be considered simultaneously [56]. When considering a two states asymmetric reaction, the potential energy of the system (E_{sys}) can be computed from the force field energies of the reactants (E_r) and products (E_p) and the mixing energy term (E_{mix}) by solving Eq. (1):

$$\begin{vmatrix} E_r - E_{sys} & E_{mix} \\ E_{mix} & E_p - E_{sys} \end{vmatrix} = 0 \quad (1)$$

Although it was initially used to simulate enzymatic reactions [57], it has also been applied to studying the reaction pathways of organic reactions such as S_N2 reactions [58,59], alkylation reactions [60] and ester cleavage [61]. However, as EVB has never been applied to asymmetric organic reactions, this method is outside the scope of this review and only a quick description is given herein. For more details, the readers are referred to excellent reviews [53,55,57]. In addition, although EVB can be exploited to compute the entire PES of reactions, we will only discuss the computation of TSs geometries and energies of a system progressing from reactants to products.

For EVB to accurately predict TS energies and structures, some factors should be considered. A first issue comes from the use of force fields. Most force fields are only meant to reproduce the heats of formation and compare relative energies of molecules with identical connectivity. Within EVB, this issue is addressed by the use of a correcting term which ensures that the relative energy of the reactants and products is accounted for. This correcting term can be computed using QM methods. Another issue arises when structures are far away from the energy minimum. Force fields have difficulties with distorted structures and some force fields (more specifically class I force fields) do not have an accurate description of the van der Waals energy term at short distance (i.e., steep Lennard Jones potential). To overcome this short-coming, more complex functions that better represent the PES for distorted structures can be used, such as the Morse-like potential in CFF [62] or the most popular MM3 [63–67]. An excellent discussion of these limitations can be found in an account by Jensen and Norrby [20].

In order to explore the PES including the TS, the EVB mapping potential (Eq. (2)) is used to drive molecular dynamic simulation over high energy barriers such a saddle point corresponding to TSs:

$$E_{model} = (1 - \lambda)E_{reactant} + \lambda E_{product} \quad (2)$$

In the case of a two state asymmetric reactions, this model would be a weighed sum of the energy of the reactants and products. These energies are then projected onto the true PES considering the mixing term (Eq. (1)). Typically λ values close to 0.5 correspond to a minimum on the model PES allowing for adequate searching of the TS (Fig. 12). The reaction force field (RFF) [68] and multi configurational molecular mechanics (MCMM) [69,70] are similar approaches [71]. These techniques have been validated by simu-

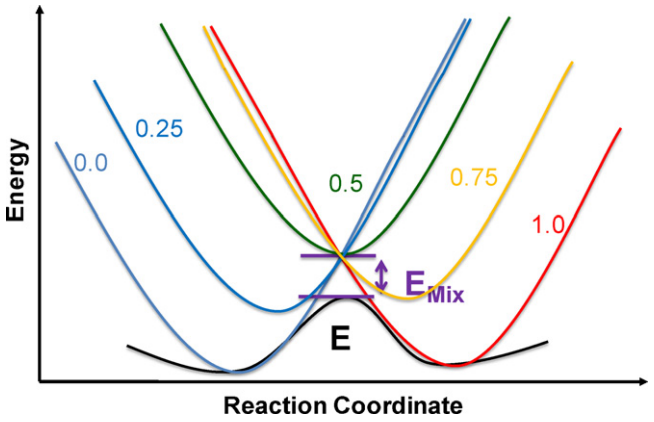

Fig. 12. All energies are calculated on the model PES then projected onto the true PES, numbers ranging from 0 to 1 indicate values of lambda.

lating reaction pathways (organic and enzymatic) but not directly applied to prediction of stereomeric excesses.

If only the relative energy between two stereoisomeric TSs is desired, it is also possible to neglect the mixing term (equivalent for both stereomeric TSs) and assume λ is equal to 0.5 for the TS. This idea has been implemented within the SEAM method [72–75]. This method has been applied to the prediction of the geometry of transition states and reactivities of reactants but not to the prediction of stereoselectivities. The initial versions of SEAM [72,73] were validated on simple reactions such as S_N2 displacement of alkyl halides. These initial studies demonstrated good correlations between experimental and predicted reactivities. In later studies on the study of achiral pericyclic reactions, Jensen and co-workers compared the SEAM predictions to a semi-empirical method (PM3) again demonstrating a good correlation [74,75].

Another option, instead of relying on force field parameters developed using QM methods as in Q2MM or mixing the energies of reactant and product is to automatically create the TS parameters on the fly as in the program Ace developed by Corbeil et al. [76]. Based on the Hammond–Leffler postulate, Ace calculates the TS energy by duplication of interactions (Fig. 13). For example, the energy of each of the forming bonds in Fig. 13 is computed as weighed sum of a non-bond and a covalent interaction. This weighed sum also allows Ace to consider the earliness or lateness of the TS. By doing so, Ace avoids going through QM methods for development of the TS parameters as in TSFF methods along with avoiding calculation of multiple energies for a single point along the PES as with EVB and SEAM methods. The major drawback of Ace is the lack of consideration of asynchronous reactions. All the forming/breaking bonds are assumed to advance simultaneously (same λ values). As a validation, Ace accurately predicted the correct stereoisomers of 41 out of 44 Diels–Alder reactions, with failures associated with carbohydrate-based chiral auxiliaries. This reveals a major limitation of FFs and their inability to predict conformations of structures on which they have not been trained. Ace was also applied to a series of 40 aldol reactions having correctly predicted the correct stereoisomer in 38 cases.

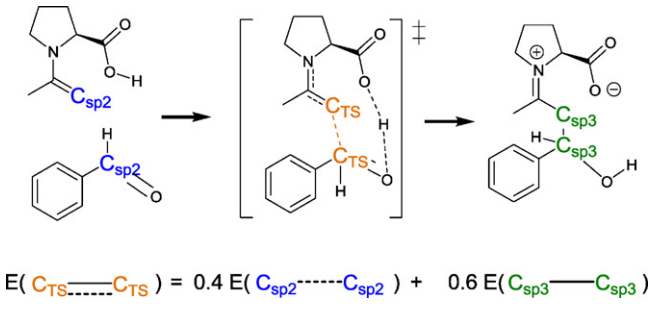

Fig. 13. TS energy for carbon–carbon bond formation for proline catalyzed aldol reaction. Figure uses λ = 0.6 in equation.

2.4. Using quantitative structure selectivity relationship (QSSR) for virtual screening of asymmetric catalysts

Both QM and MM method provide TS energies as the output, but both require that the TS structure is being modeled correctly by the method, especially in multiple TS reactions. In the field of drug design this would be similar to knowing the structure of your drug target (e.g., enzyme). When the target structure is unknown, medicinal chemists may turn to quantitative structure activity relationship (QSAR) methods which provide a relationship between the biological activity and either physical or chemical properties of a molecule. This approach is rational as in practice active compounds typically share similar chemical features and physical properties. When QSAR techniques are applied to the field of asymmetric catalyst development, they are rebranded QSSR since selectivity and not activity is the desired predicted property. QSSR is defined as the process that relates chemical structure quantitatively to a chemical transformation [77]. In essence, the simplest QSSR technique relates a series of descriptors, whether they are constitutional, topological, geometrical and physicochemical, to chemical structures [78,79] enabling prediction of stereoselectivities without prediction of TS energies. For example, Chavali et al. [80,81] used molecular indices describing electronic structures and connectivities to predict catalytic activity and toxicity. Even though this technique was not used to predict stereomeric excesses, it was a demonstration of the potential of QSSR techniques to predict chemical properties.

It is also desirable to relate structural features directly to Gibbs free energy. Based on this, Oslob et al. [82] predicted regioselectivities in palladium catalyzed allylation (Fig. 14), it was postulated that the reactivity of the terminal allyl carbon can be evaluated by a linear relationship between descriptors such as bond distance, angles and a series of dihedral angles which describe the relative position of the allyl group, the palladium atom and the ligand with regio- and/or stereoselectivity. The regioselective ratio was found to be predicted best using four descriptors: the breaking Pd–C bond distance, two dihedrals describing the in-plane distortion and displacement of the allyl group and the final, and most influential, the energy increase associated with the incoming nucleophile. These descriptors were generated from a MM2 minimized ground state structure. This energy increase was calculated by measuring the energy difference between the palladium allyl complex alone and the minimized complex in presence of the nucleophile. Overall, this techniques yielded good correlations between experimental and predicted energies with MUE of around 0.9 kcal/mol.

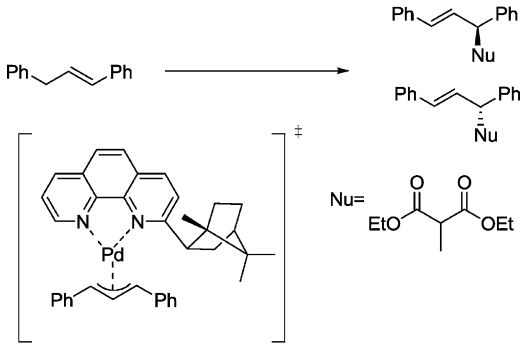

Fig. 14. Palladium catalyzed allylation.

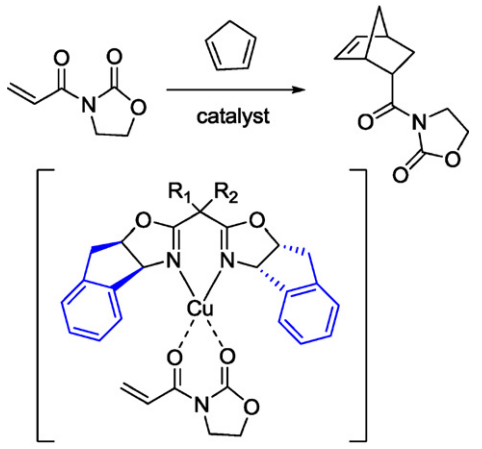

Fig. 15. bis(oxazoline)copper(II) catalyzed Diels–Alder. C5 flaps shaded in Blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

This work showed promise and the use of geometrical descriptors to relate chemical structure to regioselectivity can most likely be applied to other reactions. In the case shown in Fig. 14, regioselectivity leads to stereoselectivity as the C2-symmetrical substrate is desymmetrized.

Alvarez et al. [83] used a method termed continuous chirality measure, to determine which portions of the molecule are inducing stereoselectivity. The method was first used to rationalize the stereoselectivity of the bis(oxazoline)copper(II)-catalyzed Diels–Alder reaction. They deduced that the chirality of the adduct is mainly induced by the aromatic flaps (Fig. 15, portion in blue) which affect the orientation of the diene. Very good correlations between experimental stereoselectivities and continuous chirality measures were observed for compounds with greater than 80%ee. Upon further investigation, a new bis(oxazoline)copper(II) catalyst was proposed to be highly stereoselective (catalyst in Fig. 15) but was unfortunately not synthesized. It is worth mentioning that the use of a specifically modified MMFF94 force field to predict the stereoselectivity of this reaction was poorly accurate due to a number of factors including the twist of the copper complex and the absence of counterion in the computations. This metal complex adopts any conformation ranging from planar to tetrahedral rendering the force field development very challenging [84].

Descriptors for QSSR other than structural descriptors are quantum molecular interaction fields [85,86] (see Fig. 16). Kozlowski and co-workers superimposed optimized TS conformation of a series of known catalysts onto a grid (similar to CoMFA in the field of drug design [87]). At each point on this grid the interaction energy between the molecule under investigation and a carbon 2s electron probe is calculated using QM methods. Regression analysis is next performed on the computed grids to find regions common to all catalysts where increases in energy of the probe results in increases (green region in Fig. 16) or decreases (red region in Fig. 16) in stereoselectivities (For interpretation of the references to color in this text, the reader is referred to the web version of the article). As soon as the training with a set of known catalysts has been carried out, the grids can then be used to predict the selectivity of new catalysts. Kozlowski and co-workers applied this method to the addition of dialkylzinc to aldehydes catalyzed by β-amino alcohols and achieved excellent correlations between predicted and experimental stereoselectivities with a MUE of around 0.15 kcal/mol and 5%ee on a training set of 18 catalysts [85]. This level of accuracy was retained when a small testing set of four catalysts was used (MUE around 0.25 kcal/mol and 7%ee). Further use of this model led to the development of *trans*-1-amino-2-hydroxy cyclohexane derivatives as a chiral catalyst for amino alcohol catalyzed aldehyde alkylation reactions. In fact, this is a rare example of a real application of one of the technique to the design and synthesis of a novel asymmetric catalyst. Excellent correlation between experiments and predictions led to an easy identification of catalysts with low, medium and high asymmetric induction [88,89]. At the same time Lipkowitz et al. used quantum molecular interaction fields on bis(oxazoline)copper(II) catalyzed Diels–Alder reactions (Fig. 15). Their model identified key regions where addition or removal of steric bulk could increase selectivity and yield excellent correlations between predicted and experimental stereoselectivities. It is also possible to omit the alignment step and use a distance correlation map as in Sciabola et al. [90]. Their belief that the regression analysis is dependent on the alignment method led to the develop-

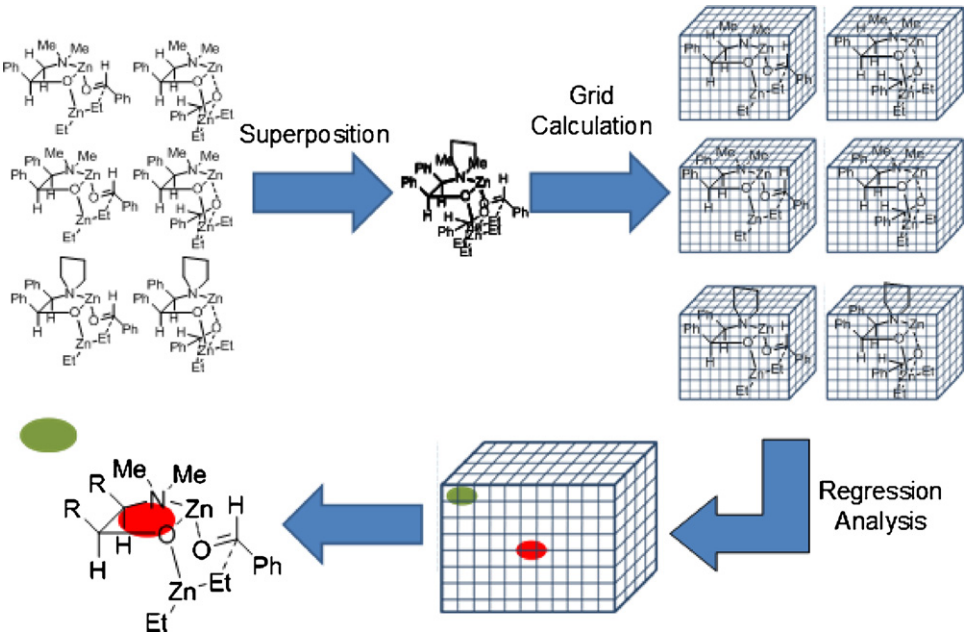

Fig. 16. QSSR using quantum mechanical interaction field analysis in the design of chiral amino alcohols for dialkylzinc addition to aldehydes.

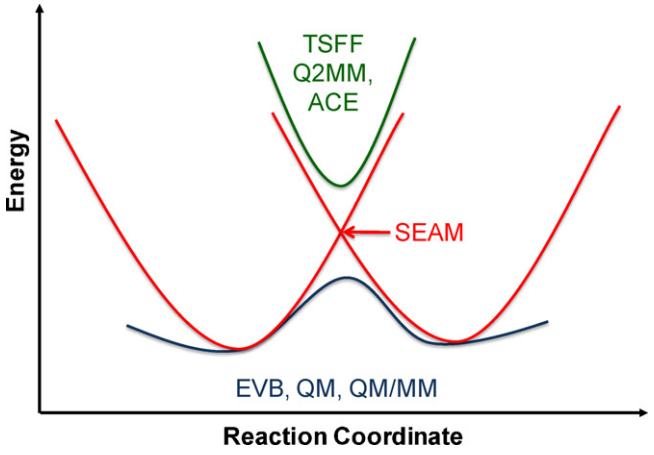

Fig. 17. Summary of methods used to find transition states.

ment of this model. Application to both the addition of dialkyl zinc to aldehydes and bis(oxazoline)copper(II) catalyzed Diels–Alder reaction yielded models with accuracy similar to the Kozlowski method [85].

3. Summary

In summary, it is possible to effectively search for the TS of reactions using a plethora of methods (Fig. 17). However caution is needed when selecting the method. For a complete one time search of a PES, either QM or EVB methods are recommended. For highly accurate prediction of stereomeric excesses, QM methods should also be used. If one wants to perform a VS or computer-aided optimization of a catalyst, specialized MM methods such as TSFF, EVB, SEAM or ACE are more suited.

One must also take results published with a grain of salt. Method published with low errors in prediction of %ee may be a result of selecting only reactions/cases where the %ee is very high. This would result in cases where the prediction accuracy may be due to the non-linearity in error in %ee with respect to error in TS energy (Fig. 1). Therefore a better measure of accuracy would be the use of MUE in prediction of TS energies.

Acknowledgments

We thank NSERC and Virochem Pharma for financial support as well as the Canadian Foundation for Innovation for financial support through the New Opportunities Fund program. This is also a National Research Council of Canada publication, number NRC 50687.

References

[1] A.B. Richon, Drug Discov. Today 13 (2008) 665–669.
[2] A.B. Richon, Drug Discov. Today 13 (2008) 659–664.
[3] R.V.C. Guido, G. Oliva, A.D. Andricopulo, Curr. Med. Chem. 15 (2008) 37–46.
[4] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, C.R. Corbeil, Br. J. Pharmacol. 153 (2008) S7–S26.
[5] B.K. Shoichet, S.L. McGovern, B. Wei, J.J. Irwin, Molecular Informatics: Confronting Complexity, May 13–16th 2002.
[6] C.N. Parker, J. Biomol. Screen 10 (2005) 647–648.
[7] P.T. Lang, I.D. Kuntz, G.M. Maggiora, J. Bajorath, J. Biomol. Screen 10 (2005) 649–652.
[8] C.R. Corbeil, E. Therrien, N. Moitessier, Curr. Comput. Aided Drug Des. 5 (2009) 241–263.
[9] B.J. Lynch, D.G. Truhlar, J. Phys. Chem. A 105 (2001) 2936–2941.
[10] P.J. Donoghue, P. Helquist, O. Wiest, J. Org. Chem. 72 (2007) 839–847.
[11] S.T. Schneebeli, M.L. Hall, R. Breslow, R. Friesner, J. Am. Chem. Soc. 131 (2009) 3965–3973.
[12] S. Bahmanyar, K.N. Houk, H.J. Martin, B. List, J. Am. Chem. Soc. 125 (2003) 2475–2479.
[13] P.H.Y. Cheong, K.N. Houk, Synthesis (2005) 1533–1537.
[14] C.B. Shinisha, R.B. Sunoj, Org. Biomol. Chem. 5 (2007) 1287–1294.
[15] D.H. Ess, G.O. Jones, K.N. Houk, Adv. Synth. Catal. 348 (2006) 2337–2361.
[16] P.H.Y. Cheong, H. Zhang, R. Thayumanavan, F. Tanaka, K.N. Houk, C.F. Barbas Iii, Org. Lett. 8 (2006) 811–814.
[17] S. Mitsumori, H. Zhang, P.H.Y. Cheong, K.N. Houk, F. Tanaka, C.F. Barbas, J. Am. Chem. Soc. 128 (2006) 1040–1041.
[18] J.M. Brown, R.J. Deeth, Angew. Chem. Int. Ed. 48 (2009) 4476–4479.
[19] K.N. Houk, P.H.Y. Cheong, Nature 455 (2008) 309–313.
[20] F. Jensen, P.O. Norrby, Theor. Chem. Acc. 109 (2003) 1–7.
[21] K.N. Houk, M.N. Paddon-Row, N.G. Rondan, Science 231 (1986) 1108–1117.
[22] H. Lin, D.G. Truhlar, Theor. Chem. Acc. 117 (2007) 185–199.
[23] D. Balcells, F. Maseras, New J. Chem. 31 (2007) 333–343.
[24] G. Drudis-Solé, G. Ujaque, F. Maseras, A. Lledós, Chem. Eur. J. 11 (2005) 1017–1029.
[25] G. Ujaque, F. Maseras, A. Lleds, J. Am. Chem. Soc. 121 (1999) 1317–1323.
[26] K.N. Houk, N.G. Rondan, Y.D. Wu, J.T. Metz, M.N. Paddon-Row, Tetrahedron 40 (1984) 2257–2274.
[27] N. Moitessier, B. Maigret, F. Chrétien, Y. Chapleur, Eur. J. Org. Chem. (2000) 995–1005.
[28] N. Moitessier, C. Henry, C. Len, Y. Chapleur, J. Org. Chem. 67 (2002) 7275–7282.
[29] A.J. DelMonte, J. Haller, K.N. Houk, K.B. Sharpless, D.A. Singleton, T. Strassner, A.A. Thomas, J. Am. Chem. Soc. 119 (1997) 9907–9908.
[30] D.J. Harriman, G. Deslongchamps, J. Comput. Aided Mol. Des. 18 (2004) 303–308.
[31] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, J. Comput. Chem. 19 (1998) 1639–1662.
[32] J.D. Harriman, G. Deslongchamps, J. Mol. Model. 12 (2006) 793–797.
[33] D.J. Harriman, G.F. Deleavey, A. Lambropoulos, G. Deslongchamps, Tetrahedron 63 (2007) 13032–13038.
[34] D.J. Harriman, A. Lambropoulos, G. Deslongchamps, Tetrahedron Lett. 48 (2007) 689–692.
[35] J.E. Eksterowicz, K.N. Houk, Chem. Rev. 93 (1993) 2439–2461.
[36] A.C.T. Van Duin, S. Dasgupta, F. Lorant, W.A. Goddard Iii, J. Phys. Chem. A 105 (2001) 9396–9409.
[37] K.D. Nielson, A.C.T. Van Duin, J. Oxgaard, W.Q. Deng, W.A. Goddard Iii, J. Phys. Chem. A 109 (2005) 493–499.
[38] K. Chenoweth, A.C.T. Van Duin, P. Persson, M.J. Cheng, J. Oxgaard, W.A. Goddard Iii, J. Phys. Chem. C 112 (2008) 14645–14654.
[39] P.O. Norrby, J. Mol. Struct. Theochem. 506 (2000) 9–16.
[40] T. Rasmussen, P.O. Norrby, J. Am. Chem. Soc. 125 (2003) 5130–5138.
[41] P.J. Donoghue, E. Kieken, P. Helquist, O. Wiest, Adv. Synth. Catal. 349 (2007) 2647–2654.
[42] P. Rydberg, L. Olsen, P.O. Norrby, U. Ryde, J. Chem. Theor. Comput. 3 (2007) 1765–1773.
[43] H. Becker, P.T. Ho, H.C. Kolb, S. Loren, P.O. Norrby, K.B. Sharpless, Tetrahedron Lett. 35 (1994) 7315–7318.
[44] P. Fristrup, G.H. Jensen, M.L.N. Andersen, D. Tanner, P.O. Norrby, J. Organomet. Chem. 691 (2006) 2182–2198.
[45] P. Fristrup, D. Tanner, P.O. Norrby, Chirality 15 (2003) 360–368.
[46] P.O. Norrby, T. Rasmussen, J. Haller, T. Strassner, K.N. Houk, J. Am. Chem. Soc. 121 (1999) 10186–10192.
[47] P.O. Norrby, P. Brandt, T. Rein, J. Org. Chem. 64 (1999) 5845–5852.
[48] P.J. Donoghue, P. Helquist, P.-O. Norrby, O. Wiest, J. Chem. Theor. Comput. 4 (2008) 1313–1323.
[49] P.J. Donoghue, P. Helquist, P.-O. Norrby, O. Wiest, J. Am. Chem. Soc. 131 (2009) 410–411.
[50] C.R. Landis, P. Hilfenhaus, S. Feldgus, J. Am. Chem. Soc. 121 (1999) 8741–8754.
[51] S. Feldgus, C.R. Landis, J. Am. Chem. Soc. 122 (2000) 12714–12727.
[52] S. Feldgus, C.R. Landis, Organometallics 20 (2001) 2374–2386.
[53] A. Warshel, R.M. Weiss, J. Am. Chem. Soc. 102 (1980) 6218–6226.
[54] J. Aqvist, A. Warshel, Chem. Rev. 93 (1993) 2523–2544.
[55] S.C.L. Kamerlin, A. Warshel, Faraday Discuss. 145 (2010) 71–106.
[56] K. Kolmodin, J. Aqvist, Int. J. Quantum Chem. 73 (1999) 147–159.
[57] A. Warshel, Proc. Natl. Acad. Sci. U.S.A. 81 (1984) 444–448.
[58] J.K. Hwang, G. King, S. Creighton, A. Warshel, J. Am. Chem. Soc. 110 (1988) 5297–5311.
[59] I. Benjamin, J. Chem. Phys. 129 (2008).
[60] J. Villà, J. Bentzien, A. González-Lafont, J.M. Lluch, J. Bertran, A. Warshel, J. Comput. Chem. 21 (2000) 607–625.
[61] V. Luzhkov, J. Aqvist, J. Am. Chem. Soc. 120 (1998) 6131–6137.
[62] A. Warshel, J. Chem. Phys. 55 (1971) 3327–3335.
[63] N.L. Allinger, Y.H. Yuh, J.-H. Lii, J. Am. Chem. Soc. 111 (1989) 8551–8566.
[64] J.-H. Lii, N.L. Allinger, J. Am. Chem. Soc. 111 (1989) 8566–8575.
[65] J.-H. Lii, N.L. Allinger, J. Am. Chem. Soc. 111 (1989) 8576–8582.
[66] N.L. Allinger, F. Li, L. Yan, J. Comput. Chem. 11 (1990) 848–867.
[67] N.L. Allinger, F. Li, L. Yan, J.C. Tai, J. Comput. Chem. 11 (1990) 868–895.
[68] A.K. Rappé, M.A. Pietsch, D.C. Wiser, J.R. Hart, L.M. Bormann-Rochotte, W.M. Skiff, Mol. Eng. 7 (1997) 385–400.
[69] Y. Kim, J.C. Corchado, J. Villa, J. Xing, D.G. Truhlar, J. Chem. Phys. 112 (2000) 2718–2735.
[70] D.G. Truhlar, J. Comput. Chem. 28 (2007) 73–86.
[71] S.C.L. Kamerlin, J. Cao, E. Rosta, A. Warshel, J. Phys. Chem. B 113 (2009) 10905–10915.
[72] F. Jensen, J. Am. Chem. Soc. 114 (1992) 1596–1603.

[73] F. Jensen, J. Comput. Chem. 15 (1994) 1199–1216.
[74] F. Jensen, J. Chem. Phys. 119 (2003) 8804–8808.
[75] P.T. Olsen, F. Jensen, J. Chem. Phys. 118 (2003) 3523–3531.
[76] C.R. Corbeil, S. Thielges, J.A. Schwartzentruber, N. Moitessier, Angew. Chem. Int. Ed. 47 (2008) 2635–2638.
[77] P. Gedeck, R.A. Lewis, Curr. Opin. Drug Discov. Dev. 11 (2008) 569–575.
[78] M. Hoogenraad, G.M. Klaus, N. Elders, S.M. Hooijschuur, B. McKay, A.A. Smith, E.W.P. Damen, Tetrahedron Asymmetry 15 (2004) 519–523.
[79] J.B. Van Der Linden, E.J. Ras, S.M. Hooijschuur, G.M. Klaus, N.T. Luchters, P. Dani, G. Verspui, A.A. Smith, E.W.P. Damen, B. McKay, M. Hoogenraad, QSAR Comb. Sci. 24 (2005) 94–98.
[80] S. Chavali, B. Lin, D.C. Miller, K.V. Camarda, Comput. Chem. Eng. 28 (2004) 605–611.
[81] B. Lin, S. Chavali, K. Camarda, D.C. Miller, Comput. Chem. Eng. 29 (2005) 337–347.
[82] J.D. Oslob, B. AÌŠkermark, P. Helquist, P.O. Norrby, Organometallics 16 (1997) 3015–3021.
[83] S. Alvarez, S. Schefzick, K. Lipkowitz, D. Avnir, Chem. Eur. J. 9 (2003) 5832–5837.
[84] R.J. Deeth, N. Fey, Organometallics 23 (2004) 1042–1054.
[85] M.C. Kozlowski, S.L. Dixon, M. Panda, G. Lauri, J. Am. Chem. Soc. 125 (2003) 6614–6615.
[86] K.B. Lipkowitz, M. Pradhan, J. Org. Chem. 68 (2003) 4648–4656.
[87] R.D. Cramer, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[88] J.C. Ianni, V. Annamalai, P.-W. Phuan, M. Panda, M.C. Kozlowski, Angew. Chem. Int. Ed. 45 (2006) 5502–5505.
[89] J. Huang, J.C. Ianni, J.E. Antoline, R.P. Hsung, M.C. Kozlowski, Org. Lett. 8 (2006) 1565–1568.
[90] S. Sciabola, A. Alex, P.D. Higginson, J.C. Mitchell, M.J. Snowden, I. Morao, J. Org. Chem. 70 (2005) 9025–9027.