ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

Prediction of Stereochemistry using Q2MM

Published as part of the Accounts of Chemical Research special issue "Computational Catalysis for Organic Synthesis".

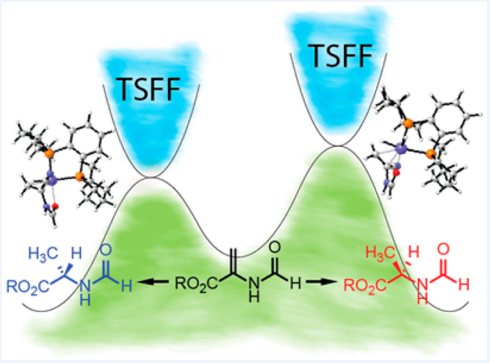
Eric Hansen,[†] Anthony R. Rosales,[†] Brandon Tutkowski,[†] Per-Ola Norrby,[‡] and Olaf Wiest*[,†,§]

[†]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, United States

[‡]Pharmaceutical Technology and Development, AstraZeneca, Pepparedsleden 1, SE-431 83 Mölndal, Sweden

[§]Lab of Computational Chemistry and Drug Design, School of Chemical Biology and Biotechnology, Peking University, Shenzhen 518055, China

CONSPECTUS: The standard method of screening ligands for selectivity in asymmetric, transition metal-catalyzed reactions requires experimental testing of hundreds of ligands from ligand libraries. This "trial and error" process is costly in terms of time as well as resources and, in general, is scientifically and intellectually unsatisfying as it reveals little about the underlying mechanism behind the selectivity. The accurate computational prediction of stereoselectivity in enantioselective catalysis requires adequate conformational sampling of the selectivity-determining transition state but has to be fast enough to compete with experimental screening techniques to be useful for the synthetic chemist. Although electronic structure calculations are accurate and general, they are too slow to allow for sampling or fast screening of ligand libraries. The combined requirements can be fulfilled by using appropriately fitted transition state force fields (TSFFs) that represent the transition state as a minimum and allow fast conformational sampling using Monte Carlo.



Quantum-guided molecular mechanics (Q2MM) is an automated force field parametrization method that generates accurate, reaction-specific TSFFs by fitting the functional form of an arbitrary force field using only electronic structure calculations by minimization of an objective function. A key feature that distinguishes the Q2MM method from many other automated parametrization procedures is the use of the Hessian matrix in addition to geometric parameters and relative energies. This alleviates the known problems of overfitting of TSFFs. After validation of the TSFF by comparison to electronic structure results for a test set and available experimental data, the stereoselectivity of a reaction can be calculated by summation over the Boltzman-averaged relative energies of the conformations leading to the different stereoisomers.

The Q2MM method has been applied successfully to perform virtual ligand screens on a range of transition metal-catalyzed reactions that are important from both an industrial and an academic perspective. In this Account, we provide an overview of the continued improvement of the prediction of stereochemistry using Q2MM-derived TSFFs using four examples from different stages of development: (i) Pd-catalyzed allylation, (ii) OsO_4-catalyzed asymmetric dihydroxylation (AD) of alkenes, (iii) Rh-catalyzed hydrogenation of enamides, and (iv) Ru-catalyzed hydrogenation of ketones. In the current form, correlation coefficients of 0.8−0.9 between calculated and experimental ee values are typical for a wide range of substrate−ligand combinations, and suitable ligands can be predicted for a given substrate with ∼80% accuracy. Although the generation of a TSFF requires an initial effort and will therefore be most useful for widely used reactions that require frequent screening campaigns, the method allows for a rapid virtual screen of large ligand libraries to focus experimental efforts on the most promising substrate−ligand combinations.

INTRODUCTION

Asymmetric catalysis is a major focus of synthetic organic chemistry. The experimental discovery of enantioselective catalysts typically involves trial-and-error approaches, thus necessitating many hours of bench work or specialized machinery to partially automate the process. This process provides only indirect insight into the origin of ligand performance and contributes little to the rational design of novel ligands. Computational approaches to enantioselective catalysis can engage in a productive interplay with experimental observations, making significant contributions by providing structural and mechanistic data rationalizing experimental data and quantitatively accurate predictions of selectivity before experimental ligand screening.

The accurate calculation of $\Delta G^{\ddagger}$ is a daunting task. In contrast, the selectivity of a reaction depends only on the relative rates between the pathways leading to the different products, e.g., different enantiomers, as indicated by the solid red and blue pathways in Figure 1. Thus, only $\Delta\Delta G^{\ddagger}$ at the transition states is

Received: January 20, 2016
Published: April 11, 2016

DOI: 10.1021/acs.accounts.6b00037
Acc. Chem. Res. 2016, 49, 996−1005

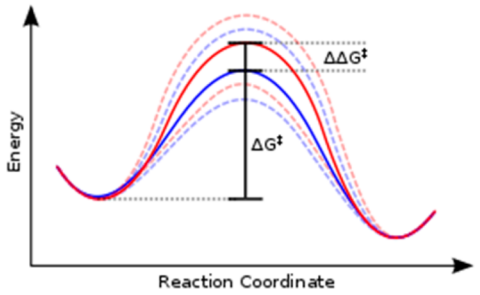


Figure 1. Diastereomeric transition states leading to enantiomeric products.

needed to predict selectivity, leading to significant error cancellation if the two relevant transition states are very similar, e.g., for diastereomeric transition states leading to enantiomeric products. The remainder of this account will focus on this case, and detailed analyses of each case can be found in the references cited.

Multiple transition states (TSs), e.g., different conformations of the TSs, can lead to the same enantiomer (dotted lines in Figure 1). These can be close enough in energy to the lowest energy TS to be relevant or even lower if the initial conformation chosen is not the lowest energy TS. In this case, the observed ee is determined by the Boltzmann distributions for each diastereomeric TS as shown in eq 1

$$\%ee = \frac{er-1}{er+1} \times 100\% \text{ with } er = \frac{\sum_{i\in R} e^{-\Delta\Delta G_i^{\ddagger}/RT}}{\sum_{j\in S} e^{-\Delta\Delta G_j^{\ddagger}/RT}} \quad (1)$$

where i and j are summed over all the conformations for each diastereomeric TS. In practice, all $\Delta\Delta G^{\ddagger}$ are taken relative to an arbitrary reference energy, generally the lowest one for each diastereomer. Historically, only single conformations from limited manual conformational searches or educated guesses were considered in computational selectivity predictions. As the systems studied computationally approached the experimentally relevant ones, the number of conformations to be considered became unfeasible for quantum mechanics (QM) methods. This is particularly problematic for systems involving large, flexible ligands where it was shown that the experimentally observed selectivities cannot be explained by a single TS.[1] Furthermore, for predictions to be useful, they must be faster than experimental screens for possibly hundreds of ligands while still being quantitatively accurate. Despite significant advances in the speed of QM and QM/MM methods, such simulations are still intractable.

A possible alternative to the problem of rapid conformational sampling is the use of fast molecular mechanics (MM) or force field (FF) methods. Although these are more commonly associated with the calculation of ground states, their use for the description of TSs has a long history. One of the first uses of an empirical FF was to describe the rotational barriers in biphenyls.[2] As early as 1929, the reaction of H and H_2 was modeled by mixing the potential energy functions of the ground states.[3] This concept was generalized as the valence bond (VB) method,[4] which gave rise to a large number of related methods,[5,6] such as multiconfigurational MM (MCMM),[7] RFF,[8] SEAM,[9] ACE,[10] and most prominently empirical VB (EVB).[11] These methods share the common element of treating the transition state by an appropriate mix of the reactant and product PES as described by a classical FF but use different approaches to determine the interaction of two ground-state diabatic PES at the transition state, as shown in Figure 2.

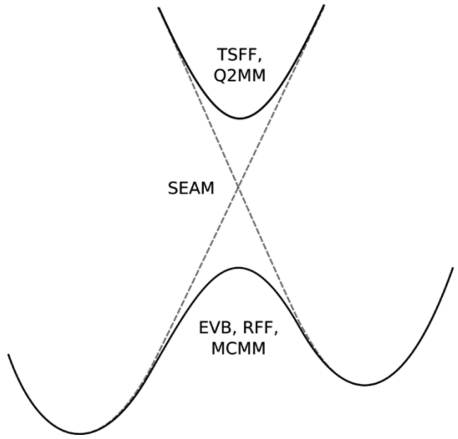


Figure 2. Comparison of force field methods for TS modeling.

An alternative method is to model the TS not as an energetic maximum but as a minimum using transition state force fields (TSFFs). TSFFs have a 50-year history of explaining and predicting relative rates and selectivities in organic reactions.[12] Conceptually, TSFFs should be more accurate in describing the TS compared to an approximation by geometrically and electronically different ground state FFs, which are frequently only capable of describing the system at small distortions from the equilibrium structure. However, a TSFF necessitates the generation of new parameters specific to the TS of interest, and the PES of the reactant, product, and TS are discontinuous. Although the second point is less relevant to computational studies of selectivity as discussed above, a practical method for the prediction of stereochemistry requires a fast and ideally automated method for the generation of reaction-specific TSFFs.

THE QUANTUM-GUIDED MOLECULAR MECHANICS (Q2MM) METHOD

Over the last two decades, we developed the Q2MM[13] method as a fast and accurate method for the parametrization of FFs. The basic idea of the Q2MM method is the automated fitting of the FF parameters using only electronic structure calculations, as described in Figure 3, by minimizing the objective function shown in eq 2.

A prerequisite for the application of the Q2MM method is knowledge of the selectivity-determining step of a given reaction. The first step is then to determine reference data for a training set consisting of small model TSs. These small model TSs should cover the minimum structures needed for capturing the essential steric and electronic features of the reaction in the definition of the new parameters close to the reaction center. They should also avoid generating "noise" from the parts of the structure that will be represented by transferable ground state parameters from the underlying force field. For each of these TSs, a full geometry optimization and frequency calculation using an appropriate electronic structure method and standard quantum mechanical code is performed, which also yields the partial charges of the atoms.

The second step is to decide for which atoms to generate new parameters and which functional form of the force field to use. Sometimes several different possibilities need to be fitted to obtain the best balance of capturing all relevant interactions with

DOI: 10.1021/acs.accounts.6b00037
Acc. Chem. Res. 2016, 49, 996−1005

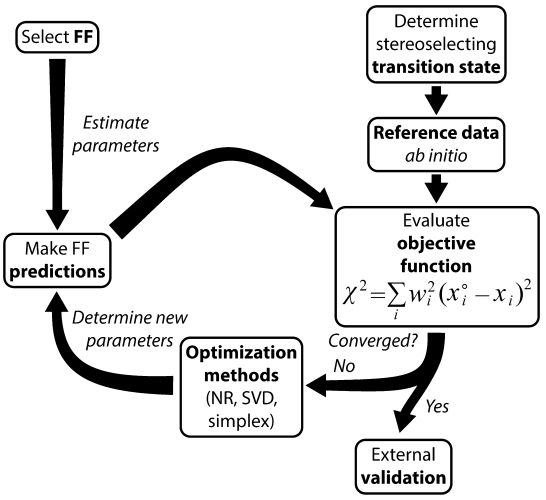


Figure 3. Flowchart of the Q2MM method.

a minimum number of parameters. Most current applications of Q2MM concern organometallic catalysts for which no ground state FF parameters are available. Thus, new atom types must be added to describe the metals, and all FF parameters involving these new atom types must be fitted. As the majority of the atoms in the system will be described by standard FF parameters, high-quality parameters have to be available for the systems to be studied. For the small organic molecules under discussion here, the MM3* force field[14] was shown to be suitable, but Q2MM can be used to fit parameters for any force field.

In the third step, the FF parameters are iteratively fit to reproduce the QM reference data by minimization of the objective function

$$\chi^2 = \sum_i w_i^2 (x_i^\circ - x_i)^2 \quad (2)$$

where x_i° is the QM-derived reference data point, x_i is the corresponding FF calculated data point, and w_i is the weight. The weight is typically set to be the inverse of the acceptable error for a given data type (e.g., bond lengths).[15] Thus, a converged FF with N data points in its training set should converge to a penalty function value no greater than N. A combination of gradient-based and simplex optimization techniques is used to minimize the penalty function.[13] We have automated the fitting procedures outlined in Figure 3 into code that works with a number of QM and MM programs, which is freely available from the authors.

Although all bonded and nonbonded parameters of the force field need to be fit, we found it useful to start with the electrostatic parameters alone. Coulombic parameters are fit to reproduce RESP or CHELPG charges calculated using standard software packages. As Q2MM is designed to work closely with MM3*, existing van der Waals parameters can be used.

A key difference between Q2MM and most traditional manual or automated methods for fitting system-specific FF parameters[16−19] is the use of the Hessian Matrix for the fitting of force constants of bonded parameters together with geometric data for reference values in the Q2MM method.[13,20,21] The Hessian matrix describes how the energy changes with respect to geometric distortions, which is important for a number of reasons. The Hessian matrix in Cartesian coordinates consists of $3N \times 3N$ data points, where N is the number of atoms. This provides an abundance of reference data for FF optimizations, thus circumventing the previously criticized overfitting of parameters in TSFFs.[22] It also contains the information for an accurate description of the energetic penalty upon distortion from the equilibrium geometry that is expected to be relevant for bulky ligands often used in stereoselective catalysis. Finally, it provides a convenient means to invert the first order saddle point obtained from the QM calculations (Figure 4, left) to the minimum used in TSFFs (Figure 4, right).[23] Inverting the curvature of the PES along the reaction coordinate causes a geometry optimization to head toward the TS structure, allowing the use of standard geometry optimization techniques included in MM packages to locate TSs. Flipping this curvature is possible by decomposing the Hessian into its corresponding eigenvectors V and eigenvalues S

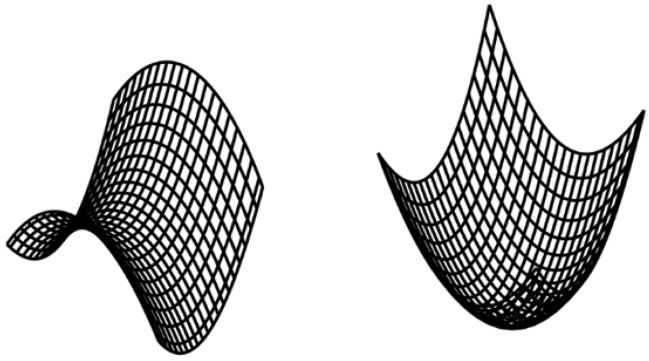


Figure 4. Representation of the Hessian's description of the PES about a TS (left) and minimum (right).

$$H = VSV^T \quad (3)$$

and replacing the negative eigenvalue of the TS with a large positive value or fitting directly to the eigenmodes and eigenvectors while ignoring the eigenmode with negative eigenvalues.[24]

Bonded parameters are optimized following optimization of the nonbonded parameters. The reference, or "ideal", bond and angle values are initially set to average values obtained from the training set and often change little throughout the optimization. Bond and angle force constants are initially set to fairly high standard values and are mainly influenced by the Hessian data during parametrization. Dihedral force constants are often fit to reproduce energies from QM torsional scans. However, this data may not be readily obtainable at the TS depending on the size and coordination of the system. Another option is to optimize dihedral force constants such that the MM FF reproduces the QM-derived Hessian and potentially the relative energies of different conformers from the training set. Fitting the dihedral parameters is typically one of the more problematic phases of the parametrization, as the choice of periodicity requires some expertise. In general, this should be done after an initial fitting of other parameters; torsional parameters can to some extent be seen as error corrections due to insufficient representation of through-space interactions in the force field but should not be allowed to take over the modeling of interactions that are more properly described by bond, angle, or nonbonded parameters. The same is true to an even larger extent for any cross-terms included in the force field.[24]

The final step of the Q2MM method is validation of the TSFF by comparing the results for geometries, Hessian matrix values, and relative energies from optimizations using the TSFF for transition states obtained from electronic structure calculations

DOI: 10.1021/acs.accounts.6b00037
Acc. Chem. Res. 2016, 49, 996−1005

that were not part of the training set. Once satisfactory agreement has been achieved, the TSFF can be used to run Monte Carlo (MC) conformational searches for the TS, leading to each of the stereoisomers as shown in Figure 1. The relative energies of the resulting conformations can be Boltzmann averaged to predict stereoselectivities using eq 2. After validation by comparison to experimentally known stereoselectivities, this can be used to rapidly predict ee's and screen suitable ligands for a given catalyst/substrate combination.

APPLICATIONS OF Q2MM TO STEREOSELECTIVE CATALYSIS

Palladium-Mediated Allylation

The first applications of the predecessor of the Q2MM method to the prediction of stereoselectivity were qualitative predictions of palladium-catalyzed allylation (Figure 5) using ground state

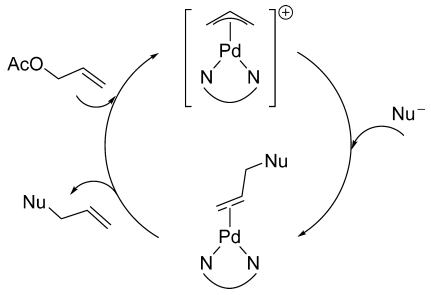


Figure 5. Catalytic cycle for the palladium-mediated allylation of nucleophiles.

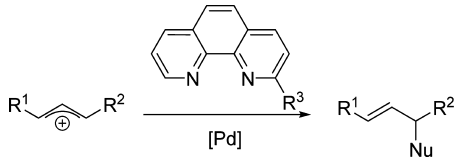
FFs.[25,26] Several chiral bidentate ligands had been shown to be effective for enantioselective and regioselective catalysis of this reaction,[25] and MM parameters for metal−ligand interactions including palladium−nitrogen and η^3-allylpalladium complexes were under development.[27] However, several competing reaction pathways with different selectivity determining steps dependent on the different substrates and ligands were thought to be operating, prompting two studies using ground state FFs for substituted phenanthroline ligands and several substrates to evaluate selectivity.

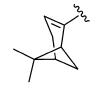
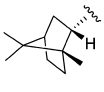
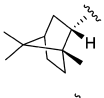
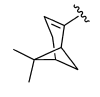
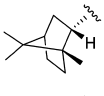
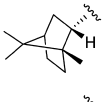
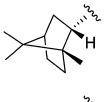
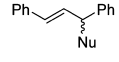
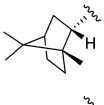
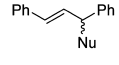
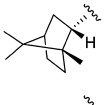
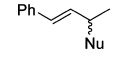
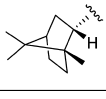
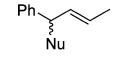
The first study to explore the enantioselectivity of several substrates with chiral phenanthroline ligands is summarized in Table 1. The results rank the ligands in a qualitative fashion as high, low, or modest in enantioselectivity.[25] Although these results contributed to the development of the Q2MM method, these predictions had similar reliability to a chemist's intuition of steric and electronic effects.

The second study combined MM with linear free energy relationships (LFER) between experimentally observed selectivities ($\Delta\Delta G^{\ddagger}$) and the ground state parameters of the allyl-Pd structures to improve the accuracy of the predictions.[26] Several achiral phenanthroline-derived ligands were included in the validation of the resulting FF to explore regioselectivity. This work emphasized steric interactions and thus included ligands and substrates with little variation in electronic structure. The focus on diethyl-methylmalonate and DMF cancelled effects of the nucleophile and solvent, respectively. The resulting predictions (Table 2) were better than the earlier qualitative predictions. The inclusion of the LFER emphasized the need to examine TSs.

These early studies provided several important insights into the further development of the Q2MM method. First, reactions

Table 1. Qualitative Predictions Using Only a Ground State FF



Entry	R^3	Nu	Product	Expt'l Ratio	Predicted Ratio
1		Diethyl methylmalonate		1	Low
2		Diethyl methylmalonate		1.06	Low
3		Dimethyl malonate		1.63	Low
4		Diethyl methylmalonate		1.27	Low
5		Diethyl methylmalonate		1.17	Low
6		Dimethyl malonate		1.38	Modest
7		Diethyl methylmalonate		3.44	Modest
8		Dimethyl malonate		9.0	High
9		Diethyl methylmalonate		2.23	High
10		Dimethyl malonate		1.99	Modest
11		Dimethyl malonate		39	High

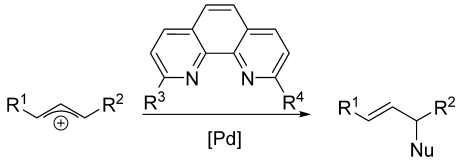
with a well-defined rate- and selectivity-determining step are more suitable for reaction-specific FFs. Utilizing only general parameters of a ground state FF will usually provide only qualitative predictions at best. The performance of the combined MM and LFER, although accurate, requires several experiments with a range of selectivities for parametrization.

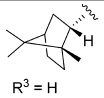
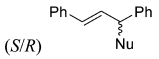
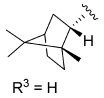
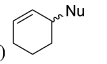
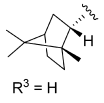
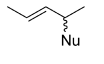
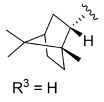
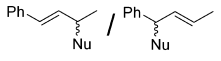
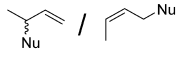
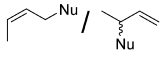
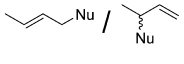
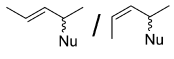
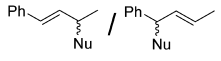
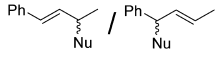
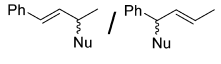
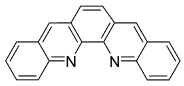
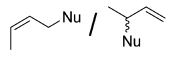
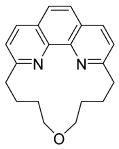
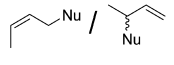
Osmium Tetroxide-Catalyzed Asymmetric Dihydroxylation (AD) of Alkenes

The first iterations of the Q2MM method attempted to address these problems by using the results from QM calculations for the parametrization of a TSFF. The osmium tetroxide AD reaction of alkenes (Figure 6)[28−30] has a wide scope, mild conditions, high yields, and excellent enantioselectivity. The intense debate over a concerted [3 + 2] or stepwise [2 + 2] mechanism was resolved in favor of the former by a combination of isotope effect experiments and DFT calculations.[31]

DOI: 10.1021/acs.accounts.6b00037
Acc. Chem. Res. 2016, 49, 996−1005

Table 2. Combined Ground State FF and LFER Energy Comparison (kJ/mol) of Nucleophilic Addition Using Diethyl Methylmalonate



Entry	Ligand	Products	Expt'l $\Delta\Delta G^{\ddagger}$	Calc'd $\Delta\Delta G^{\ddagger}$
1	R^4 =  R^3 = H	 (*S*/*R*)	1.98	0.56
2	R^4 =  R^3 = H	 (*R*/*S*)	3.10	2.02
3	R^4 =  R^3 = H	 (*S*/*R*)	0.60	1.56
4	R^4 =  R^3 = H		2.06	1.74
5	R^3 = Me, R^4 = *t*Bu		0.00	-0.05
6	R^3 = R^4 = Me		1.00	-0.47
7	R^3 = R^4 = Me		11.4	11.4
8	R^3 = R^4 = Me		9.64	8.92
9	R^3 = R^4 = Me		5.73	5.18
10	R^3 = H, R^4 = Me		0.92	1.05
11	R^3 = H, R^4 = *t*Bu		3.73	6.01
12			2.10	-0.42
13			2.94	3.46

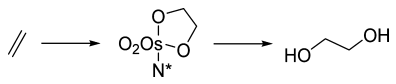


Figure 6. Osmium Tetroxide-Catalyzed Asymmetric Dihydroxylation (AD) of Alkenes.

The training set included several TS structures with simple model ligands of ammonia or trimethylamine along with various alkene substrates (Figure 7). Validation of the results from the Q2MM-derived TSFF to QM calculations as well as experimental results (Figure 8) show excellent agreement with an ~2.5 kJ/mol mean unsigned error (MUE), which is considerably better than the "chemical accuracy" target of 4 kJ/mol. The TSFF was able to predict the correct major enantiomer, and in most cases, the predicted ee is close to the experimental values. Substrates where predictions deviated significantly from the experimental results typically show low ee's (indicating low energy differences and hence large errors) and are often those that react slowly, possibly due to alternative pathways (e.g., nonselective background reactions without the chiral ligand).

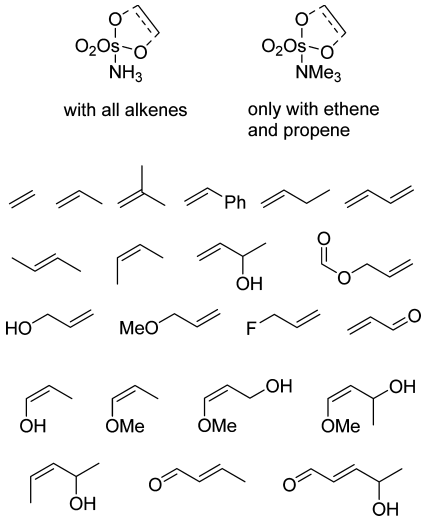


Figure 7. Osmium complexes and alkenes used as the QM training set.

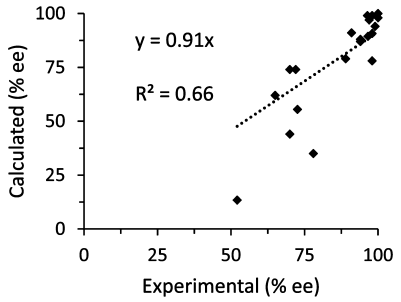


Figure 8. Selectivity comparison of Q2MM-derived FF and experimental results for AD reaction.

The Q2MM-derived TSFF, together with additional experimental studies, allowed for the refinement of the original mnemonic for face-selectivity based on DHQD- or DHQ-derived ligands (Figure 9, top).[28,29] Analysis of the lowest energy conformations from the conformational searches using the TSFF provided an improved mnemonic (Figure 9, bottom) that gives better insights into the repulsive and attractive interactions responsible for stereoselectivity. Additionally, the TSFF also rationalizes the low stereoselectivity for certain substrates that cannot be described by the simple mnemonic.

The AD reaction is also one of the few cases for which a direct comparison can be made between Q2MM and QM/MM methods.[32] Both methods were found to have similar accuracies, and each has its own strengths and weaknesses. Careful selection of the structures to be calculated in the QM/MM studies deemphasized the need for a large conformational search, but the number of cases that could be studied is still limited. Conversely, the effort needed for the generation of the Q2MM TSFF is considerable, but once available, it is orders of magnitude faster than the QM/MM calculations, enabling extensive sampling and/or high throughput virtual screening of substrates and

DOI: 10.1021/acs.accounts.6b00037
Acc. Chem. Res. 2016, 49, 996−1005

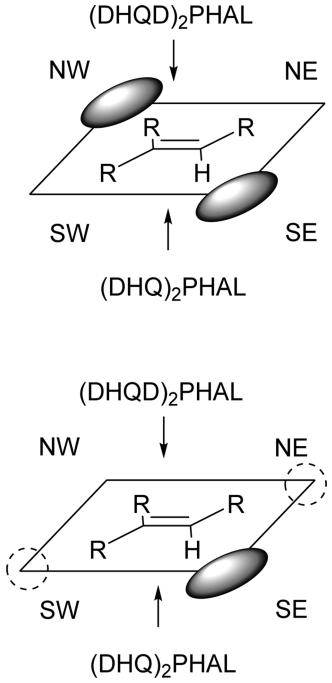


Figure 9. (top) Original mnemonic to predict the selectivity of the Sharpless AD using $(DHQD)_2PHAL$ and $(DHQ)_2PHAL$ ligands. (bottom) Revised mnemonic derived from analysis with the Q2MM TSFF. Ovals and dotted circles represent steric bulk and attractive areas of the ligand, respectively.

ligands. Both the QM/MM and Q2MM approaches provided solutions to a well-known problem, the lack of dispersive attractive forces in DFT calculations. In both cases, the regions subjected to QM studies are small enough so that dispersion interactions could be ignored, whereas the attractive vdW interactions that are an important feature of the dihydroxylation reaction were captured using the accurate vdW potential in MM3.

Rhodium-Catalyzed Hydrogenation of Enamides

The rhodium-catalyzed hydrogenation of enamides is a reaction of great synthetic importance to both academia and industry for the synthesis of natural and unnatural amino acids.[33,34] A generalized example is shown in Figure 10. The reaction is

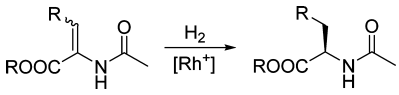


Figure 10. Rh-Catalyzed hydrogenation of enamides.

known to be highly sensitive to both ligand and substrate structure, necessitating time- and resource-intensive experimental screens to identify highly enantioselective combinations. The reaction is therefore a prime candidate for the development of a Q2MM TSFF, which would allow a fast virtual screening of ligands to identify promising candidates among the more than 200 commercially available bisphosphines used for this reaction.

The mechanism of the reaction had been elucidated previously[35−38] and is summarized in Figure 11, but the stereoselecting step was not unambiguously identified. A distinctive feature of this mechanism is the "anti-lock-and-key" principle, meaning that the major square planar conformation frequently leads to the minor product enantiomer.[35,39,40] The

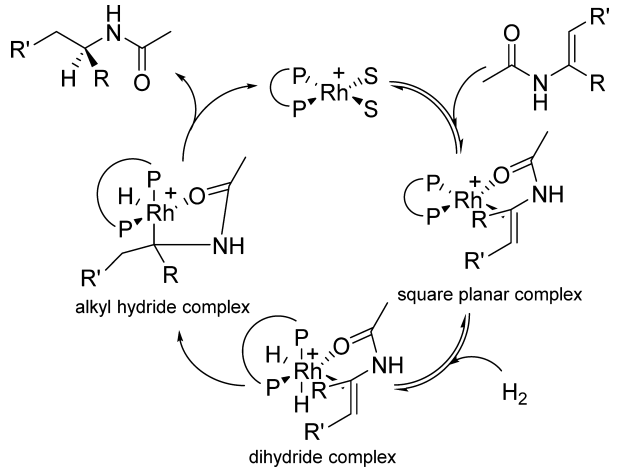


Figure 11. Overall mechanism of rhodium-catalyzed hydrogenation of enamides.

observed enantioselectivity is based on the reactivity of the substrate−catalyst complexes with hydrogen to generate the product, emphasizing the need for an analysis of the competing TSs.

On the basis of DFT results for the stereodetermining step of the mechanism,[41] we developed a Q2MM TSFF.[42] The QM training set consisted of four structurally different chiral and achiral bidentate phosphine ligands (ZDMP, DMPE, (*R*,*R*)-MeDuPHOS, and (*R*,*R*)-Me-BPE) and the substrate *α*-formamidoacrylonitrile, Figure 12, which was used previously in computational studies of the reaction.[36−38]

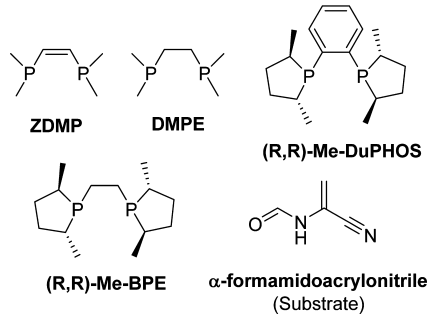


Figure 12. Ligands and substrate used in the QM training set.

In total, nine structures were optimized at the B3LYP/lacvp** level of theory for use as the training set. Using the functional form of MM3*, new atom types were added, and the necessary parameters were optimized using the automated procedure outlined in Figures 3 and 4 to reproduce the electronic and geometric structures of the TS calculated by DFT. A comparison of the QM and MM data verified that the developed TSFF could accurately reproduce the training set.[42] After the initial validation toward the training set, 13 ligands and seven substrates were chosen for a virtual screening.[43] The ligands **A**−**M** and substrates **1**−**7**, shown in Figure 13, were chosen to represent structural and electronic diversity, including different types of chirality, substrate geometries, and different protecting groups, and for the availability of experimental results under similar conditions in the literature.[43]

The correlation of the Q2MM ee predictions with the known experimental ee has an R^2 of 0.90 as shown in Figure 14,

DOI: 10.1021/acs.accounts.6b00037
Acc. Chem. Res. 2016, 49, 996−1005

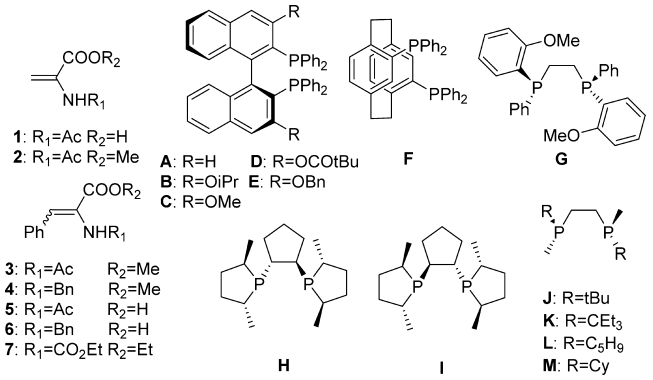


Figure 13. Ligands **A**–**M** and substrates **1**–**7** used in virtual screening and selectivity comparison of Q2MM-derived FF and experimental results for Rh-catalyzed hydrogenation of enamides.

indicating that the FF was correctly modeling the reaction in the vast majority of cases.

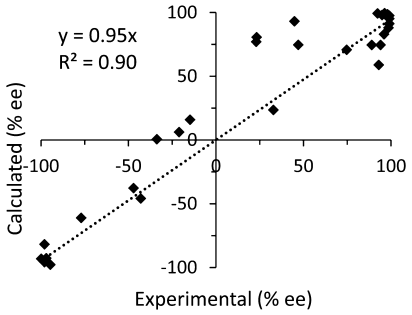


Figure 14. Selectivity comparison of Q2MM-derived FF and experimental results for Rh-catalyzed hydrogenation of enamides.

Most importantly, highly selective substrate/ligand combinations are rapidly identified with an ∼80% accuracy, e.g., of five selected ligands, four will actually give a high selectivity for a given substrate. This accuracy, which is much higher than in virtual screens used in medicinal chemistry, provides the experimental chemist with a rapid tool to focus efforts on a very small number of ligands for a given substrate.

There are two types of outliers. As was the case in the earlier studies, the largest deviations occur for reactions with low stereoselectivity due to the small energy values involved. As these cases are unlikely to be of interest to the experimental chemist, this is acceptable. There are a small number of false positives, i.e., ligands that are predicted to be highly selective but are experimentally known to have low selectivity. Interestingly, these consistently involve the bulkiest ligands tested, such as **B**, **E**, and **G**, suggesting a possible dissociation of the ligand or substrate.[44]

Ruthenium-Catalyzed Hydrogenation of Ketones

After several methodological improvements to the Q2MM procedure, it was applied to the ruthenium-catalyzed hydrogenation reaction of ketones, which occurs readily using the Noyori catalyst (Figure 15).[45] The selectivity and rate-determining transition state was elucidated previously and is shown in Figure 16.[46] Hydrogenation of the ketone is concerted with hydrogen delivery mediated by the ruthenium and a dative nitrogen. The reaction is attractive for application of the Q2MM method for a variety of reasons. First, it is synthetically important in that it has the potential to be both chemo- and stereoselective

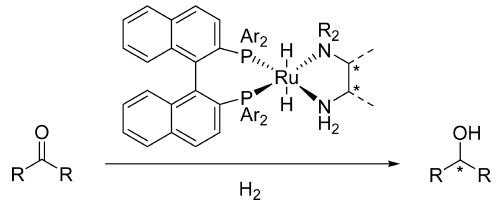


Figure 15. Hydrogenation of a ketone with a Noyori catalyst.

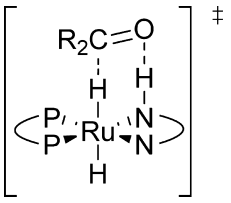


Figure 16. TS depicting a reparameterized reaction center.

in the reduction of ketones to alcohols. Chiral alcohols are a common motif in natural products as well as drugs and drug-like molecules. Although there are many ways to selectively reduce a ketone, very few provide a catalytic approach that uses inexpensive hydrogen gas. Second, several mechanistic studies have elucidated a single step that determines stereoselectivity. Third, the combinatorics of two different chiral ligands leads to a number of possible combinations too large to be explored experimentally. The use of the Q2MM method to perform a high throughput screen has the potential to focus an experimental screen on manageable numbers while exploring a wide chemical space.

The QM training set for parametrization included several combinations of the diamine ligands and substrates with a common phosphine ligand as is shown in Figure 17. All

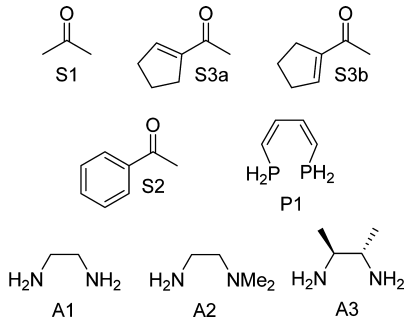


Figure 17. Substrates and ligands used for the QM training set.

calculations were completed using the B3LYP/lacvp* level of theory. The TS structures were used to parametrize the MM3* FF for atoms involved in the reaction center shown in Figure 16. The internal validation toward the QM data demonstrated the success of the automated parametrization for the reproduction of geometries and vibrational modes.

Although the training set used simplified ligands and substrates, the independent test set, calculated using the same QM method, included more complex systems such as BINAP-type and chiral diamine ligands. The conformational space of these more complex ligands was explored using the TSFF and located conformations that were subjected to QM calculations at the same level of theory. The stereoselectivity based on the Boltzmann averaging was compared to the one obtained by using only the lowest energy conformers of the two diasteromers.

DOI: 10.1021/acs.accounts.6b00037
Acc. Chem. Res. 2016, 49, 996–1005

Comparison of both $\Delta\Delta E^{\ddagger}$ methods produces similar predictions, but the MUE of the latter was significantly larger than using the Boltzmann averaging, highlighting the importance of higher energy conformations.

The validation by comparison to experimental data gave good overall agreement between experimental and calculated ee with an MUE of only 2.7 kJ/mol (Figure 18), showcasing the

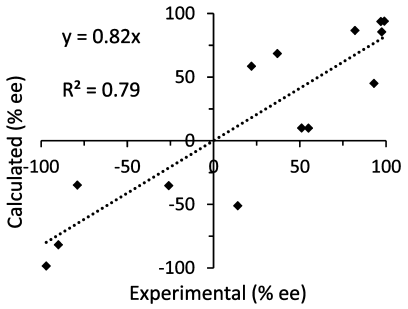


Figure 18. Selectivity comparison of Q2MM-derived FF and experimental results for Ru-catalyzed hydrogenations of ketones.

improvements made to the Q2MM method in this more complex reaction. A further inspection of energy contributions of a ligand−substrate combination can assist in elucidating the major contribution of selectivity. One such example would be the calculation of the (*S*)-TolBINAP/(*R*)-DMAPEN/(*E*)-chalcone complex. Most of the energy differences between the two diastereomeric TSs are derived from vdW forces, especially with respect to edge−face interactions of the aromatic rings. This highlights that many ligand−substrate combinations have unique interactions that cannot be predicted by general rules.

CONCLUSIONS

For the past two decades, the Q2MM method has been developed and successfully applied to perform the virtual screening of ligands in enantioselective reactions.[47,48] Although the parametrization of the reaction-specific TSFF requires an upfront investment to obtain the underlying QM data and fit and validate the force field, the speed and accuracy of the computational predictions make Q2MM-derived TSFFs an attractive complement to experimental screens of substrate−ligand combinations for commonly used transition metal-catalyzed transformations. To this end, the Q2MM code and the validated force fields are freely available to the community via github.com/Q2MM/q2mm. The ultimate goal of the method is to enable bench chemists and other scientists who may not have specific experience in computational methods to easily use Q2MM to make quick, useful predictions that greatly decrease the number of experiments necessary for the identification of suitable ligands, thus increasing the efficiency of experimental approaches. The incorporation of the existing TSFF discussed here as well as additional ones into easy-to-use graphical interfaces with predefined ligand libraries can make a predictive virtual ligand screen an integral part of protocol development in enantioselective catalysis.

The concepts outlined in this Account are, in principle, also transferrable to other types of selectivity (such as regioselectivity) and other problems where computationally expensive calculations have to be repeated many times to achieve adequate sampling, such as QM/MM molecular dynamics on enzymes. Force field methods cannot generally be used to calculate energies of isomers that differ in which parameters are used, such as regioisomers, but the force field could be used to explore the conformational space of each isomer (within an isoparametric set), and the energy difference in nonisoparametric comparisons could be calibrated using a few DFT calculations or by comparison to selected experimental data.[23] Current developments of Q2MM will investigate the applicability of the methodology to these problems.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Olaf.G.Wiest.1@nd.edu.

Notes

The authors declare no competing financial interest.

Biographies

Eric Hansen received his B.S. in Chemistry from Grand Valley State University, Michigan. He is currently a graduate student at the University of Notre Dame working with Professor Wiest.

Anthony Rosales is a native of Montana, where he received a B.A. with honors in Chemistry from Carroll College in Helena. As a second year graduate student at the University of Notre Dame du Lac, he serves the needs of Professors Olaf Wiest and Paul Helquist.

Brandon Tutkowski received his B.S. in Chemistry in 2013 from Concordia University Wisconsin in Mequon. He is currently a graduate student at the University of Notre Dame working in the combined research group of Professors Paul Helquist and Olaf Wiest.

Per-Ola Norrby received his Ph. D. from the Royal Institute of Technology in Stockholm, Sweden. Until 2013, he was professor in organic synthesis at the University of Gothenburg, but then moved to AstraZeneca as a principal scientist. He is an adjunct professor at the Universities of Gothenburg and Notre Dame.

Olaf Wiest received his Ph.D. from the University of Bonn, Germany, and was a Feodor Lynen Postdoctoral Fellow at UCLA. He is a full professor in the Department of Chemistry and Biochemistry at the University of Notre Dame and a visiting professor at PKUSZ. He is a Camille Dreyfus Teacher Scholar and a Fellow of the AAAS.

ACKNOWLEDGMENTS

We are grateful for support of Q2MM-related research by the National Science Foundation (CHE1058075 and CHE0833220), the Petroleum Research Fund (PRF#47810-AC1), the National Institutes of Health (R01GM111645), the Swedish Research Council (#2010-4856), and the FP7-funded SYNFLOW program (www.synflow.eu).

REFERENCES

(1) Hatanaka, M.; Maeda, S.; Morokuma, K. Sampling of Transition States for Predicting Diastereoselectivity Using Automated Search Methods: Aqueous Lanthanide-Catalyzed Mukaiyama Aldol Reaction. *J. Chem. Theory Comput.* **2013**, *9*, 2882−2886.

(2) Westheimer, F. H. A Calculation of the Energy of Activation for the Racemization of 2,2′-Dibromo-4,4′-Dicarboxydiphenyl. *J. Chem. Phys.* **1947**, *15*, 252.

(3) London, F. Quantenmechanische Deutung Des Vorgangs Der Aktivierung. *Z. Elektrochem. Angew. Phys. Chem.* **1929**, *35*, 552−555.

(4) Eyring, H.; Polanyi, M. Zur Berechnung Der Aktivierungswärme. *Naturwissenschaften* **1930**, *18*, 914−915.

(5) Florián, J. Comment on Molecular Mechanics for Chemical Reactions. *J. Phys. Chem. A* **2002**, *106*, 5046−5047.

(6) Truhlar, D. G. Reply to Comment on Molecular Mechanics for Chemical Reactions. *J. Phys. Chem. A* **2002**, *106*, 5048−5050.

DOI: 10.1021/acs.accounts.6b00037
Acc. Chem. Res. 2016, 49, 996−1005

(7) Kim, Y.; Corchado, J. C.; Villa, J.; Xing, J.; Truhlar, D. G. Multiconfiguration Molecular Mechanics Algorithm for Potential Energy Surfaces of Chemical Reactions. *J. Chem. Phys.* **2000**, *112*, 2718−2735.
(8) Rappé, A. K.; Pietsch, M. A.; Wiser, D. C.; Hart, J. R.; Bormann-Rochotte, L. M.; Skiff, W. M. RFF, conceptual development of a full periodic table force field for studying reaction potential surfaces. *Mol. Eng.* **1997**, *7*, 385−400.
(9) Jensen, F. Locating Minima on Seams of Intersecting Potential Energy Surfaces. An Application to Transition Structure Modeling. *J. Am. Chem. Soc.* **1992**, *114*, 1596−1603.
(10) Weill, N.; Corbeil, C. R.; De Schutter, J. W.; Moitessier, N. Toward a Computational Tool Predicting the Stereochemical Outcome of Asymmetric Reactions: Development of the Molecular Mechanics-Based Program ACE and Application to Asymmetric Epoxidation Reactions. *J. Comput. Chem.* **2011**, *32*, 2878−2889.
(11) Åqvist, J.; Warshel, A. Simulation of Enzyme Reactions Using Valence Bond Force Fields and Other Hybrid Quantum/classical Approaches. *Chem. Rev.* **1993**, *93*, 2523−2544.
(12) Eksterowicz, J. E.; Houk, K. N. Transition-State Modeling with Empirical Force Fields. *Chem. Rev.* **1993**, *93*, 2439−2461.
(13) Norrby, P.-O.; Liljefors, T. Automated Molecular Mechanics Parameterization with Simultaneous Utilization of Experimental and Quantum Mechanical Data. *J. Comput. Chem.* **1998**, *19*, 1146−1166.
(14) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.
(15) Norrby, P.-O.; Brandt, P. Deriving Force Field Parameters for Coordination Complexes. *Coord. Chem. Rev.* **2001**, *212*, 79−109.
(16) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885−1891.
(17) Huang, L.; Roux, B. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput.* **2013**, *9*, 3543−3556.
(18) Wu, J. C.; Chattree, G.; Ren, P. Automation of AMOEBA Polarizable Force Field Parameterization for Small Molecules. *Theor. Chem. Acc.* **2012**, *131*, 1−11.
(19) Vanduyfhuys, L.; Vandenbrande, S.; Verstraelen, T.; Schmid, R.; Waroquier, M.; Van Speybroeck, V. QuickFF: A Program for a Quick and Easy Derivation of Force Fields for Metal-Organic Frameworks from Ab Initio Input. *J. Comput. Chem.* **2015**, *36*, 1015−1027.
(20) Maple, J. R.; Hwang, M.-J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. Derivation of Class II Force Fields. I. Methodology and Quantum Force Field for the Alkyl Functional Group and Alkane Molecules. *J. Comput. Chem.* **1994**, *15*, 162−182.
(21) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(22) Sherrod, M. J.; Menger, F. M. "Transition-State Modeling" Does Not Always Model Transition States. *J. Am. Chem. Soc.* **1989**, *111*, 2611−2613.
(23) Norrby, P.-O.; Brandt, P.; Rein, T. Rationalization of Product Selectivities in Asymmetric Horner-Wadsworth-Emmons Reactions by Use of a New Method for Transition-State Modeling. *J. Org. Chem.* **1999**, *64*, 5845−5852.
(24) Limé, E.; Norrby, P.-O. Improving the Q2MM Method for Transition State Force Field Modeling. *J. Comput. Chem.* **2015**, *36*, 244−250.
(25) Peña-Cabrera, E.; Norrby, P.-O.; Sjögren, M.; Vitagliano, A.; De Felice, V.; Oslob, J.; Ishii, S.; O'Neill, D.; Åkermark, B.; Helquist, P. Molecular Mechanics Predictions and Experimental Testing of Asymmetric Palladium-Catalyzed Allylation Reactions Using New Chiral Phenanthroline Ligands. *J. Am. Chem. Soc.* **1996**, *118*, 4299−4313.
(26) Oslob, J. D.; Åkermark, B.; Helquist, P.; Norrby, P.-O. Steric Influences on the Selectivity in Palladium-Catalyzed Allylation. *Organometallics* **1997**, *16*, 3015−3021.
(27) Norrby, P. O.; Åkermark, B.; Hæffner, F.; Hansson, S.; Blomberg, M. Molecular Mechanics (MM2) Parameters for the (η_3-Allyl)-palladium Moiety. *J. Am. Chem. Soc.* **1993**, *115*, 4859−4867.
(28) Fristrup, P.; Jensen, G. H.; Andersen, M. L. N.; Tanner, D.; Norrby, P.-O. Combining Q2MM Modeling and Kinetic Studies for Refinement of the Osmium-Catalyzed Asymmetric Dihydroxylation (AD) Mnemonic. *J. Organomet. Chem.* **2006**, *691*, 2182−2198.
(29) Fristrup, P.; Tanner, D.; Norrby, P.-O. Updating the Asymmetric Osmium-Catalyzed Dihydroxylation (AD) Mnemonic: Q2MM Modeling and New Kinetic Measurements. *Chirality* **2003**, *15*, 360−368.
(30) Norrby, P. O.; Rasmussen, T.; Haller, J.; Strassner, T.; Houk, K. N. Rationalizing the Stereoselectivity of Osmium Tetroxide Asymmetric Dihydroxylations with Transition State Modeling Using Quantum Mechanics-Guided Molecular Mechanics. *J. Am. Chem. Soc.* **1999**, *121*, 10186−10192.
(31) DelMonte, A. J.; Haller, J.; Houk, K. N.; Sharpless, K. B.; Singleton, D. A.; Strassner, T.; Thomas, A. A. Experimental and Theoretical Kinetic Isotope Effects for Asymmetric Dihydroxylation. Evidence Supporting a Rate-Limiting "(3 + 2)" Cycloaddition. *J. Am. Chem. Soc.* **1997**, *119*, 9907−9908.
(32) Ujaque, G.; Maseras, F.; Lledós, A. Theoretical Study on the Origin of Enantioselectivity in the Bis(dihydroquinidine)-3,6-pyridazine·Osmium Tetroxide-Catalyzed Dihydroxylation of Styrene. *J. Am. Chem. Soc.* **1999**, *121*, 1317−1323.
(33) *Stereoselective Formation of Amines*; Li, W., Zhang, X., Eds.; Springer: New York, 2014.
(34) Noyori, R. Asymmetric Catalysis: Science and Opportunities (Nobel Lecture). *Angew. Chem., Int. Ed.* **2002**, *41*, 2008−2022.
(35) Halpern, J. Mechanism and Stereoselectivity of Asymmetric Hydrogenation. *Science* **1982**, *217*, 401−407.
(36) Landis, C. R.; Halpern, J. Asymmetric Hydrogenation of Methyl-(Z)-a-Acetamidocinnamate Catalyzed by {1,2-Bis((phenyl-O-anisoyl)-phosphino) ethane}rhodium(I): Kinetics, Mechanism, and Origin of Enantioselection. *J. Am. Chem. Soc.* **1987**, *109*, 1746−1754.
(37) Landis, C. R.; Brauch, T. W. Probing the Nature of H2 Activation in Catalytic Asymmetric Hydrogenation. *Inorg. Chim. Acta* **1998**, *270*, 285−297.
(38) Landis, C. R.; Hilfenhaus, P.; Feldgus, S. Structures and Reaction Pathways in Rhodium(I)-Catalyzed Hydrogenation of Enamides: A Model DFT Study. *J. Am. Chem. Soc.* **1999**, *121*, 8741−8754.
(39) Chua, P. S.; Roberts, N. K.; Bosnich, B.; Okrasinski, S. J.; Halpern, J. The Origins of the Enantioselection in Asymmetric Catalytic Hydrogenation of Amino-Acid Precursors. *J. Chem. Soc., Chem. Commun.* **1981**, 1278−1280.
(40) Schmidt, T.; Baumann, W.; Drexler, H.-J.; Arrieta, A.; Heller, D.; Buschmann, H. About the Crystal Structure of [Rh((S,S)-DIPAMP)-((Z)-2-Benzoylamino-3-(3,4-Dimethoxyphenyl)-Methyl Acrylate)]-BF4: Major or Minor Catalyst−Substrate Complex? *Organometallics* **2005**, *24*, 3842−3848.
(41) Donoghue, P. J.; Helquist, P.; Wiest, O. Ligand and Substrate Effects on the Mechanism of Rhodium-Catalyzed Hydrogenation of Enamides. *J. Org. Chem.* **2007**, *72*, 839−847.
(42) Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O. Development of a Q2MM Force Field for the Asymmetric Rhodium Catalyzed Hydrogenation of Enamides. *J. Chem. Theory Comput.* **2008**, *4*, 1313−1323.
(43) Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O. Prediction of Enantioselectivity in Rhodium Catalyzed Hydrogenations. *J. Am. Chem. Soc.* **2009**, *131*, 410−411.
(44) Gridnev, I. D.; Imamoto, T. Challenging the Major/Minor Concept in Rh-Catalyzed Asymmetric Hydrogenation. *ACS Catal.* **2015**, *5*, 2911−2915.
(45) Limé, E.; Lundholm, M. D.; Forbes, A.; Wiest, O.; Helquist, P.; Norrby, P.-O. Stereoselectivity in Asymmetric Catalysis: The Case of Ruthenium-Catalyzed Ketone Hydrogenation. *J. Chem. Theory Comput.* **2014**, *10*, 2427−2435.
(46) Dub, P. A.; Henson, N. J.; Martin, R. L.; Gordon, J. C. Unravelling the Mechanism of the Asymmetric Hydrogenation of Acetophenone by

DOI: 10.1021/acs.accounts.6b00037
Acc. Chem. Res. 2016, 49, 996−1005

[RuX2(diphosphine)(1,2-Diamine)] Catalysts. *J. Am. Chem. Soc.* **2014**, *136*, 3505−3521.
(47) Nilsson Lill, S. O.; Forbes, A.; Donoghue, P.; Verdolino, V.; Wiest, O.; Rydberg, P.; Norrby, P.-O. Application of Q2MM to Stereoselective Reactions. *Curr. Org. Chem.* **2010**, *14*, 1629−1645.
(48) Madarász, Á.; Dènes, B.; Paton, R. S. Development of a True Transition State Force Field from Quantum Mechanical Calculations. *J. Chem. Theory Comput.* **2016**, *12*, ASAP DOI: 10.1021/acs.jctc.5b01237.

DOI: 10.1021/acs.accounts.6b00037
Acc. Chem. Res. 2016, 49, 996−1005