



ChemComm

Application of Q2MM to Predictions in Stereoselective Synthesis

Journal:	<i>ChemComm</i>
Manuscript ID	CC-FEA-05-2018-003695.R1
Article Type:	Feature Article

SCHOLARONE™
Manuscripts



Chemical Communications

FEATURE ARTICLE

Application of Q2MM to Predictions in Stereoselective Synthesis

Anthony R. Rosales,^{a‡} Taylor R. Quinn,^{a‡} Jessica Wahlers,^a Anna Tomberg,^b Xin Zhang,^c Paul Helquist,^a Olaf Wiest,^{*a,c} and Per-Ola Norrby^{*a,b}

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

DOI: 10.1039/x0xx00000x

www.rsc.org/

Quantum-Guided Molecular mechanics (Q2MM) can be used to derive transition state force fields (TSFFs) that allow the fast and accurate predictions of stereoselectivity for a wide range of catalytic enantioselective reactions. The basic ideas behind the derivation of TSFFs using Q2MM are discussed and the steps involved in obtaining a TSFF using the Q2MM code, publically available at github.com/q2mm, are shown. The applicability for a range of reactions, including several non-standard applications of Q2MM, is demonstrated. Future developments of the method are also discussed.

The Importance of Stereoselective Synthesis

A basic tenet of stereochemistry is that the absolute configurations of chiral compounds have direct bearings on their interactions with other chiral compounds or systems. A critically important aspect of this principle arises in pharmaceutical chemistry. Two enantiomers of a chiral drug may have drastically different activities in a living organism due to differential interactions with chiral cellular components, such as proteins functioning as drug targets. The differences can be as great as one enantiomer of a drug having desired therapeutic effects, but the other having deleterious, toxic properties.¹

This concept has taken on increasing importance in recent years. Going back a few decades, pharmaceutical products were typically planar carbocyclic or heterocyclic compounds lacking chirality, which limited their specificity for interaction with drug targets. However, as described very effectively by others,² pharmaceutical development has now journeyed far beyond "Flatland" whereby in 2016, all of the top ten pharmaceutical products based on global sales were chiral entities, either as small molecule drugs or biologicals.³ The economic impact is enormous, amounting to \$83 billion in annual sales.

Consequently, there are well-recognized needs for methods in asymmetric synthesis that permit controlled construction of specific enantiomers of compounds.⁴ Many such methods have been found through use of trial-and-error approaches to identify reagents or catalysts that provide

products with high enantiomeric excess (e.e.).^{5,6} Experimental screening of large libraries of chiral reagents and catalysts are often conducted to identify the top performers. However, screening can be complicated by several factors such as major investments in time, supplies, and equipment, accessibility of chiral reagents and catalysts, reproducibility, and scalability.⁷ An added element of complexity is the need for entirely new catalysts to obtain satisfactory performance for many applications. The new catalysts must be designed, synthesized, and finally validated experimentally, only then to be found, in many cases, not to meet performance standards.

To mitigate these challenges, there have been considerable efforts in recent years to develop methods that can predict the stereochemical outcome of reactions. This review describes how stereoselectivities can be calculated rapidly and reliably using the quantum-guided molecular mechanics (Q2MM) method^{8,9} developed in our laboratories over the last two decades. We review different approaches to the calculation of stereoselectivity and transition states (TSs) with an emphasis on force field methods, followed by an overview of the history and philosophy of Q2MM. We then present the steps involved in the derivation of a Q2MM TSFF and a few examples of the performance of Q2MM. Finally, we provide an overview of several alternative applications of Q2MM and future directions. Successful implementation of this and other methods can minimize the amount of experimental screening that must be performed to identify satisfactory reagents or catalysts, whether previously known or newly designed to meet specific needs.

Computational Studies of Stereoselectivity

Given the importance of enantioselective synthesis in organic chemistry, it is not surprising that a wide range of experimental and computational tools have been applied to understand and ultimately predict the structural and electronic origins of stereoselection. Computational studies of this

^a Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA. E-mail: owiest@nd.edu

^b Early Product Development, Pharmaceutical Sciences, IMED Biotech Unit, AstraZeneca, Pepparedsleden 1, SE-431 83 Mölndal, Sweden. E-mail Per-Ola.Norrby@astrazeneca.com

^c Lab of Computational Chemistry and Drug Design, School of Chemical Biology and Biotechnology, Peking University, Shenzhen Graduate School, Shenzhen 518055, China.

‡ These authors contributed equally.

question can be subdivided into three conceptually different approaches that are briefly reviewed: (i) correlations of physicochemical properties with stereoselectivity, (ii) electronic structure calculations of the stereoisomeric transition structures of a reaction and (iii) studies using classical molecular mechanics force fields.

(i) Correlation methods. An early example of the first category is the “stereocartography” approach by Kozłowski and Lipkowitz.¹⁰ Here, the differences in steric and electrostatic interactions around the chiral ligands is mapped in three dimensions and statistically correlated to observed stereoselectivities. This approach is conceptually similar to the Comparative Molecular Field Analysis (CoMFA) widely used in drug discovery and development.¹¹ More recently, Sigman and coworkers have used multidimensional correlation analyses between experimentally observed stereoselectivities and a wide range of physicochemical parameters describing the steric and electronic effects of substituents to develop predictive models of stereoselectivity.¹² An advantage of these statistical approaches is that the precise mechanism and stereoselecting step of the reaction do not need to be known. Like all correlation methods, the results of these QSAR-like methods depend on the training set chosen, and extrapolation beyond the structural elements contained in the training set needs to be carefully validated.¹³ Furthermore, the lack of structural information about the stereoselecting transition state can be an impediment to the rational design of improved ligands in asymmetric catalysis.

(ii) Electronic structure calculations. The most common method to get more detailed insight into the structure of the stereocontrolling transition state(s) are explicit electronic structure calculations using wavefunction or density functional methods. This is an extremely active field with hundreds of studies reported annually. Here, only a few general trends and selected recent applications are mentioned.

The standard application of electronic structure methods is the calculation of the intermediates and transition structures involved in the catalytic cycle of an enantioselective reaction, followed by detailed analyses of the rate- and stereo-determining steps and, ideally, the formulation of experimentally verifiable hypotheses to validate the model.¹⁴ While the exponential growth of computational power and the refinement of computational methods now allows the computation of realistic model systems in a reasonable amount of time, several conceptual problems remain. Obviously, only a limited number of mechanisms can be considered in such a study, and there is always the possibility that the real pathway is not calculated. Furthermore, the increasing size of the model systems not only exponentially increases the number of possible conformers that need to be explored, but also suggests that multiple conformations of the transition states are involved in the reaction pathways leading to the two possible stereoisomeric products. If the energy differences involved are small, as is often the case for enantioselective reactions, multiple pathways need to be Boltzmann-averaged for a realistic description of the reaction.¹⁵

To achieve the acceleration necessary for the exploration of the conformational space, different strategies have been applied. One is to use QM/MM methods such as ONIOM¹⁶ that combine a QM treatment of a small number of atoms involved in the reaction with a force field treatment of the remaining molecule.¹⁷ Alternatively, the conformational search can be performed at a lower (e.g. semiempirical) level of theory and the located structures are automatically forwarded to higher-level calculations.¹⁸ Nevertheless, both methods are still too slow to allow for a comprehensive exploration of the conformational space.

(iii) Molecular mechanics force fields. An alternative approach is the treatment of the transition state using force field rather than electronic structure methods. Force field methods for transition states can be divided into general types (Figure 1).¹⁹

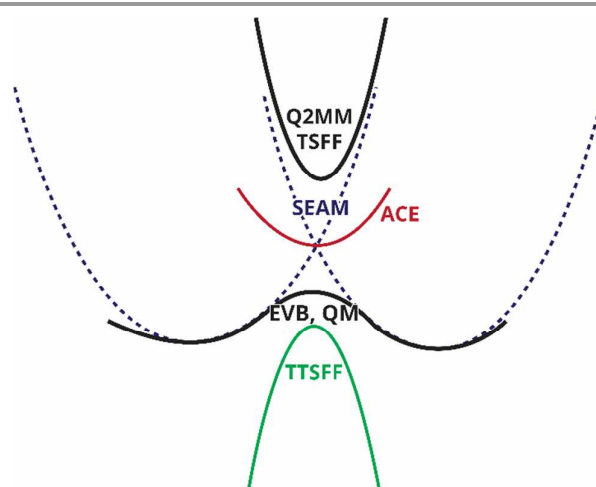


Figure 1. Force field modeling schemes for transition states

Ground state force field mixing. The first group of methods uses ground state force fields describing reactants and products and mixing them to produce a reaction potential energy surface. This type of approach was pioneered by Warshel and coworkers, who developed the empirical valence bond theory (EVB).²⁰⁻²² In EVB (Eq. 1), the reaction potential energy surface is the lowest adiabatic state obtained from a secular matrix mixing diabatic states represented by ground state force fields (E_1 and E_2 , force field energies for reactant and product). The mixing term E_{12} is a constant or a simple function of the reaction coordinate, generally determined by fitting to experiments.²³ The resulting energy surface E goes smoothly from the lowest energy of E_1 and E_2 when the difference between them is large, to a transition state, which is E_{12} below the crossing point when E_1 equals E_2 .

$$\begin{vmatrix} E_1 - E & E_{12} \\ E_{12} & E_2 - E \end{vmatrix} = 0 \Rightarrow E = \frac{E_1 + E_2 - \sqrt{(E_1 - E_2)^2 + 4E_{12}^2}}{2}$$

Eq. 1

Later approaches such as Bala's approximate VB (AVB) method,²⁴ Truhlar's multi-configuration molecular mechanics

(MCOMM) method,²⁵ the reactive force field (RFF) from Rappé,²⁶ and other related²⁷ methods differ from EVB in the nature and source of data for the mixing term E_{12} . Methods that use ground state force fields to treat transition states are easy to implement, can often apply published force field parameters, and have been widely used with considerable success. However, it is reasonable to expect that methods that are parameterized towards the transition state will perform better because they capture the essential features of the transition state more accurately.

The SEAM method by Jensen and coworkers²⁸ can be considered a special case of EVB where the coupling term is set to zero, reducing the complex problem of locating a saddle point to the more robust minimization of the energy on a seam between two hypersurfaces (Figure 1). The SEAM method delivers transition state geometries, and energies that differ from the true barrier by the constant term E_{12} , and is therefore well suited to determination of relative energies of isoparametric transition states.

Transition state force fields. The second group includes the transition state force fields (TSFF) which describe the TS as an energy minimum. A TSFF is fitted specifically to describe the transition state and therefore, unlike EVB, does not describe the entire potential energy surface with a single set of parameters. TSFF have been used for 30 years to describe relative reactivity²⁹ and kinetic isotope effects.³⁰ More systematic efforts started by Williams³¹ and Houk³² led to a number of reaction specific transition state force fields for a variety of reactions of small molecules that were manually fitted against structures, relative energies and partial charges from quantum mechanical calculations of small molecules.³³ Because a relatively small data set is used to fit a significant number of adjustable parameters in the generation of the TSFF, this method has been criticized for overfitting, most notably by Menger *et al.*, who showed that parameters obtained from a random number generator in their program FUDGIT performed similarly to a TSFF fitted against a small quantum mechanical training set.³⁴ Although this notion was rejected based on a more detailed analysis,³⁵ overfitting is a potential problem, which could lead to many possible solutions for the force field parameters that could reproduce the small training set, but do not have predictive value beyond that.

Bond-breaking force fields. A method that to some extent bridges the two classes is ReaxFF.³⁶ This approach aims at describing reactants, products, and the transition states linking them with a single force field based on ideas derived from the central force concept.³⁷ ReaxFF uses an essentially nonbonded approach to develop a general bond-order dependent potential³⁸ with parameters that depend only on the atom but not on the chemical environment. The most common applications of ReaxFF are in the areas of materials science, heterogeneous catalysis, and complex reaction mechanisms, but there have been relatively few applications to computational biophysics.³⁹

True transition state force fields. Recently, Madarász *et al.* used the principles of parameterization of transition state

force fields but allowed negative force constants for bonds forming and/or breaking in the reaction, thus producing a “true transition state force field” (TTSFF) with the correct negative curvature at the TS.⁴⁰ A TTSFF has the advantage compared to a TSFF that the shape of the TS region can be modeled accurately. The resulting force field will respond correctly to distortions along the reaction coordinate imposed by steric interactions, as opposed to the behavior in a TSFF.⁴¹ However, a significant drawback is the absence of energy minima on the PES. With a negative force constant, the energy will go toward negative infinity as the system moves away from the TS on the reaction coordinate. This will prohibit the use of common methods of conformational searching. We envision that TTSFF could be used in conjunction with a regular TSFF after a conformational search, to refine the obtained structures to a true TS.

Model potential and ACE. In conjunction with the work on EVB, Warshel and coworkers also developed a simpler *model potential*, a weighted average between the reactant and product force field: $E = (1-\lambda)E_1 + \lambda E_2$, where λ is a reaction coordinate.²⁰ The model potential will have an energy minimum that moves smoothly from reactant to product as λ is increased from 0 to 1. The original use of the model potential was to allow use of molecular dynamics (MD) at intermediate geometries, including the transition state. The energies for such intermediate geometries would then be obtained from the regular EVB expression. However, Moitessier and coworkers realized that for a specific λ , close to 0.5 for a thermoneutral reaction, the model potential would have the properties of a TSFF, and could be used similarly. This has become the basis of the ACE (asymmetric catalyst evaluation) method,⁴² which has been implemented as a general method of locating approximate transition states for reactions where reactant and product force fields are available. ACE has also been modified with TS-specific parameters to describe interactions that are absent in both product and reactant.

Historical background to Q2MM

In the late 1980's and early 1990's, we had become interested in computational methods to predict reaction stereoselectivity, particularly in the burgeoning field of asymmetric catalysis. At that time, quantum mechanical (QM) methods were greatly limited. The standard in the field was the Hartree-Fock method (HF) with a small double- ζ -valence basis set (*e.g.*, 3-21G*), which was limited in accuracy as well as in the size of systems that could be addressed with the supercomputers of the time.⁴³ For systems containing transition metals (including many of the successful chiral catalysts), more accurate methods such as MP2 or higher were needed. However, the accuracy came at an unacceptable expense; it was not feasible to do calculations on a transition metal with a ligand sphere including a dozen or more atoms.

The emerging field of Density Functional Theory (DFT) was showing significant promise in terms of scalability, but it would be another decade before such methods could be directly

applied to finding reaction barriers for systems large enough to enable prediction of stereoselectivity. More importantly, it was clear early on that HF and DFT methods had an Achilles heel that is particularly important for the field of asymmetric induction; it lacked a proper description of van der Waals (vdW) forces. Any method that only describes a static equilibrium position of electrons around the nuclei, such as HF or DFT, can only properly describe repulsive vdW forces. Attractive components, such as London dispersion, arise out of instantaneous non-equilibrium electron distributions, requiring correlated wave function theory (WFT) or equivalent approaches (Figure 2).⁴⁴

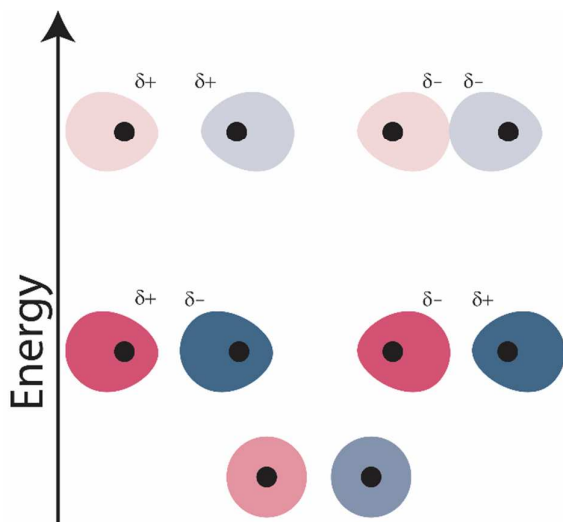


Figure 2. Dispersion. Of the possible distorted electron clouds, the attractive ones have higher population than repulsive ones, leading to a net attractive force even in systems without a ground state dipolar interaction.

Dispersion can easily be modeled by a simple analytic function with the leading term r^{-6} , where r is the distance between two atomic nuclei. By including such terms, molecular mechanics (MM) methods have long been able to describe supramolecular interactions. More recently, this type of empirical correction has been used together with DFT calculations⁴⁵, but at the time we initiated our studies, only MM methods were able to model vdW interactions for reasonably sized systems with acceptable accuracy.

In asymmetric synthesis, stereoselectivity generally is determined by the energy difference between diastereomeric transition states. The differences in barriers leading to enantiomeric products are primarily due to non-bonded interactions between substituents. Understanding and predicting stereoselectivity requires an accurate representation of vdW forces. Early experimental⁴⁶ and theoretical⁴⁷ studies of the osmium-catalyzed asymmetric dihydroxylation (AD, recognized by the Nobel Prize in 2001) highlighted the extreme influence of attractive vdW forces on both rate and selectivity in the reaction. Our earliest attempts to derive computational selectivity models, in the field of Pd-assisted allylation, focused on relative energies of high energy (η^3 -allyl)Pd intermediates modeled using ground state force

fields. It was expected that the interactions found in these models would be reflected in the selectivity-determining transition states.⁴⁸ Later efforts included TS-relevant constraints and MM-derived descriptors to make a QSAR-model for the reaction.⁴⁹ Such models allowed rationalization of the underlying causes for observed selectivities, but as with all QSAR-style models, the dependence on a training set of pre-determined experimental selectivities, and the risks inherent in predictions outside the training set motivated us to move towards fully mechanism-based models. We were inspired by the elegance of TSFF approaches (*vide supra*), but were aware of the problems in fitting numerous parameters to a small number of experimental selectivities. Houk and coworkers had shown that structural parameters (bonds and angles) could be estimated from QM-calculated model systems.³³ The group of Hagler had demonstrated how force constants could be determined using QM-calculated Hessian data (*i.e.*, the matrix of second derivatives of the energy with respect to Cartesian coordinates, calculated for vibrational analysis).⁵⁰ However, it was not *a priori* clear that Hessian data could be utilized to derive a TSFF. The local curvature is fully determined by the Hessian, and at a TS, a QM Hessian will, by definition, contain a negative curvature in the direction of the reaction coordinate. Using such a Hessian in parameterization of a force field will by necessity result in a saddle point with at least one negative force constant (a “true transition state force field” (TTSFF),⁴⁰ *vide supra*), and any attempt to use such a force field in a standard conformational search with energy minimization will diverge. The inherent convergence problems in the required transition state optimization will thus make conformational searches with current methodologies intractable. The potential energy surface of a TSFF must be positive definite at the stationary point to allow full conformational sampling. Our solution to this problem was to modify the QM data to make the Hessian positive definite. This is easily accomplished by diagonalization (as done automatically by standard QM software in a vibrational analysis), replacement of the single negative eigenvalue with a large positive value, and reformation of a modified Hessian, which will now be positive definite.⁵¹ A similar procedure had earlier been used by Dasgupta and Goddard to improve the Hessian by replacing eigenvalues derived at a low level of theory with experimental values (Figure 3).⁵²

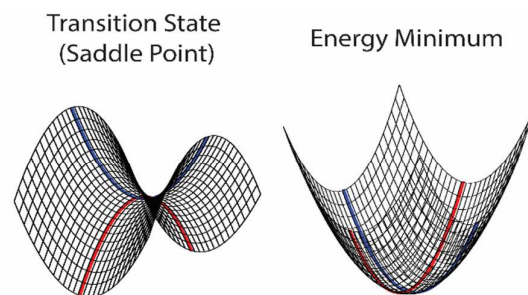


Figure 3. Eigenmode modification to transform a saddle point into an energy minimum where the reacting coordinate (red) is modified to a positive curvature while the curvature of all other coordinates (blue) are retained.

Using the modified Hessian as reference data, and in-house optimization routines to ensure the closest possible fit in a force field model,⁵³ the force field will now be positive definite and will represent the TS as an energy minimum, allowing application of standard methods for conformational searching. There will now be sufficient data from small QM model systems to allow a unique determination of each required new parameter for the TSFF. This technique marks the starting point for our Q2MM method, where a TSFF is derived solely from QM model calculations and thus can be applied in a purely predictive fashion to selectivities in asymmetric synthesis. More recently, we have introduced an alternative technique to avoid the modification to an arbitrary eigenvalue for the reaction coordinate. Instead, we extract all eigenvectors with positive eigenvalues from the diagonalization of the QM Hessian (excluding the reaction coordinate, the 3 translations and 3 rotations), and multiply these onto the MM Hessian. This produces a matrix of size $3N-7$ (where N is the number of atoms), which is compared to the diagonal matrix of QM eigenvalues. Deviations between the two matrices are summed into an objective function to be minimized during parameterization.⁵⁴ This will produce a positive definite force field, now with fewer artifacts from the enforced positive eigenvalue (the force constants can now achieve “natural” values), but the softer mode can instead introduce undesired distortions along the reaction coordinate.⁴¹

Deriving Q2MM force fields

Parameterization guided by quantum data, as implied in the name Q2MM, allows for a truly predictive method, where the training set is obtained from electronic structure calculations. To develop a TSFF, a thorough mechanistic study is needed to determine the TS of interest (Figure 4). This typically requires DFT calculations of TSs that are supported by experiments.

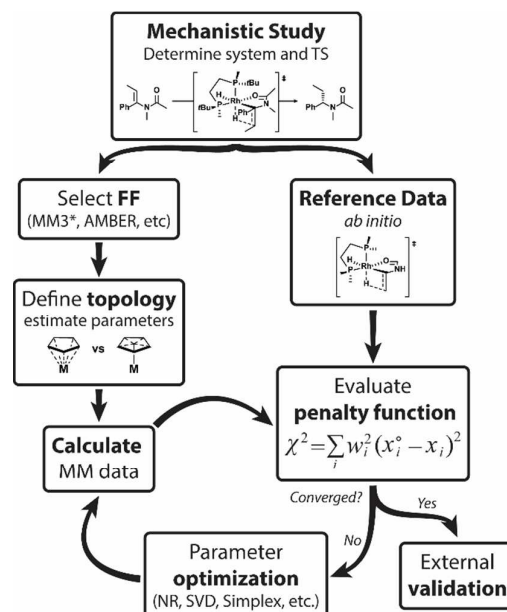


Figure 4. Flow chart of the steps taken in the Q2MM method.

Model System Selection

Once a mechanism is accepted to be operative and a TS geometry is determined, the training set can be developed from small model systems. A general principle is to start from a validated TS for an experimentally relevant system in the mechanistic study, and reduce the size to the smallest possible analog that will capture the geometrical features of the TS. In the example of the Ru-catalyzed ketone hydrogenation (Figure 5), the atoms colored in green are electronically far from the TS and are not influenced by bond breaking/forming so they are not needed in a small model system. The simplest possible model would use formaldehyde for the carbonyl, two NH_3 ligands to represent the diamine, and two PH_3 ligands for the diphosphine. However, the small model must be inspected for interactions that would not be present in a “real” system. For example, if the PH_3 moiety acts as a hydrogen bond donor in the calculation, it might be necessary to replace it with PMe_3 instead. Likewise, if the formaldehyde “tilts” in the TS, it might not be a reasonable representation of ketones, and acetone should instead be used. When the smallest possible system has been defined and validated, it is necessary to expand the training set to include larger systems in order for all parameters needed in the final force field to be represented. This is done one subunit at a time as not to introduce DFT dispersion problems in the final force field (*vide supra*). In the current example, medium sized models with different diamines, phosphines, or ketone substituents were used, but only one at a time. It is necessary at each step to control for the introduction of spurious interactions.

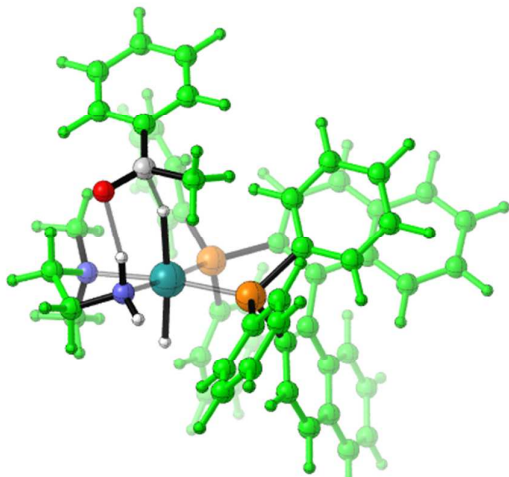


Figure 5. Transition State of the Ru-catalyzed ketone hydrogenation

Data types from QM calculations

All of the data used in the parameterization are obtained from QM calculations of the model systems. These data include: energies, structural information, electrostatic potential, and the Hessian. The large amount of data allows for an overdetermined system which resolves the earlier criticism of overfitting. Data are collected by standard DFT methods using programs such as Gaussian⁵⁵ or Jaguar.⁵⁶ Heavier atoms, like transition metals, typically require a basis set that includes effective core potentials such as Lanl2DZ. All other atoms can use a standard valence double- ζ basis set, such as 6-31g*. Although vdW interactions are limited in the model systems, empirical dispersion corrections can be used to account for relevant interactions.

Relative energies can play an important role since they are dependent on every parameter. It is important to stress the use of relative energies, since only relative energies of isoparametric systems are comparable (e.g. diastereomeric TSs) in FFs.

Model systems must be optimized to TSs yielding structural information such as bond lengths, angles, and torsions. The ability for a TSFF to reproduce the geometry at the TS is necessary to accurately describe distortions due to steric interactions.

Nonbonded interactions are the main contribution in differentiating between stereoisomers. For this reason, electrostatics play an important role. Most force fields describe electrostatics as monopoles (atomic point charges), or dipoles along bonds. Therefore partial charge calculations on the model systems are needed. Due to the default parameters within a force field, users should be aware which charge schemes reproduce similar charges. For example, the ChelpG fitting procedure produces systematically similar charges to those used in the MM3* FF. It is important to keep in mind the limitations of different charge schemes. For example, the ChelpG procedure fits partial charges to the electrostatic potential (ESP) at the surface of the structure. As

a result, atoms buried inside the structure may end up with nonsensical values.

Importantly, the Q2MM method is distinct from other parameterization methods because of the treatment of the Hessian. Reasonably, these should be heavily weighted to ensure proper description of the local TS. The Hessian typically accounts for the largest portion of available data during parameterization. As previously discussed, two methods can be used to fit these data: a MM Hessian fitted to a modified QM Hessian, or a MM eigenmatrix (diagonalized from MM Hessian and QM eigenvectors) fitted to an unmodified QM eigenmatrix.

Force field basics

Force fields were developed to quickly compute various properties of molecules by the summation of individual interactions within the molecule. The description of the individual interactions is broken down into the functional form and various parameters (e.g. force constants). The functional form of the interaction also affects the computational speed and accuracy. For example, AMBER force fields have been designed for molecular dynamics of biomolecules, where bonding interactions can be described by harmonic potentials (Eq. 2).⁵⁷ Alternatively, the force fields developed by Allinger (e.g. MM3⁵⁸) are accurate in describing energetics and dynamics of small organic molecules and use a Taylor expansion of Eq. 2 to approximate the anharmonicity of bond interactions. These two families of force fields and others have been parameterized to reproduce experimental data and therefore are intended to describe systems similar to the ones used as training sets. In the case of AMBER, parameters have been optimized to physical properties of biomolecules such as crystal structures.⁵⁷ Allinger's force fields have used spectroscopic data and heats of formation.⁵⁸ These subtleties play an important role in the selection of force field type in the development of a TSFF.

$$E = \frac{1}{2}k(r_0 - r)^2$$

Eq. 2

During the parameterization process, it is important to understand how interactions are described by the various parameters and how these parameters affect the reproduction of the reference data. The overall potential energy of a molecule can be broken down into the bonded and nonbonded interactions.

Eq. 2 or a Taylor expansion of Eq. 2 can be used to describe bond stretching and bending. The equilibrium bond length and angles are represented by r_0 , the value at which the potential energy of the interaction is at a global minimum. Additionally, some force fields will use bond dipoles to determine point charges, which can be dependent on the equilibrium bond length. The force constants, k , relate the displacement away from the equilibrium value to the potential energy. In well behaved cases, the equilibrium values relate directly to the geometries of the reference data, and the force constants can

be described from the 1,2- and 1,3-interactions from the Hessian of the reference data. Reproduction of geometries can become linearly dependent on both force constants and equilibrium values if force constants are relatively low.

Force field terms describing bond angles can pose specific problems that needs to be considered. First, transition metals and TS-geometries often contain coordination geometries that require differentiation of angles. For example, ML_6 contains two types of L-M-L angles and therefore requires a means to differentiate between the *cis*- or *trans*-interaction. Different atom types can be used for coordination complexes where ligands can be differentiated (i.e. axial and equatorial). Another way is to use geometry-dependent parameters, which have been implemented in MacroModel.⁵⁹ Second, angles around a common central atom can be linearly dependent. If this is the case, several combinations of angle parameters that will enable reproduction of the geometry as well as vibrations are needed. As an example, the H-C-H angle in methane has an angle of 109.47°, and is easily reproduced in MM by using that value as an equilibrium bond angle with an appropriate force constant to reproduce vibrations. However, it can be equally well reproduced using an equilibrium angle of 180° and a much smaller force constant. Methods to address this issue during an optimization are discussed below, but in general the parameter set that produces the lowest strain energy is often more general and therefore transferable.

The potential energy as a function of the rotation of a bond (i.e. a torsion) is periodic in nature and therefore described with a Fourier series as seen in equation (Eq. 3), where the V terms are amplitudes for the angle ω . These terms are closely related to force constants and affect reproduction of geometries and 1,4-interactions in the Hessian.

$$E_{torsion} = \frac{1}{2}V_1[1 + \cos(\omega)] + \frac{1}{2}V_2[1 - \cos(2\omega)] + \frac{1}{2}V_3[1 + \cos(3\omega)]$$

Eq. 3

The nonbonded pairwise interactions are comprised of electrostatic and van der Waals (vdW) interactions. The simplest way to describe vdW interactions, used by many force fields, is the Lennard-Jones equation (Eq. 4), where r is the pairwise distance between the atoms, ϵ is the depth of the potential, and σ is related to the equilibrium distance. More accurate force fields like MM3 use a form that is less repulsive at short range.⁵⁸ As for the electrostatic interactions, these are described from classical interactions of electric moments (e.g. monopoles, dipoles, etc.). Additionally, some modern force fields include polarizability terms to allow more dynamic electrostatic interactions. Both electrostatic and vdW interactions affect reproduction of 1,5- and higher interactions, and in some force fields softened potentials are used with 1,4-interactions. In addition to the Hessian, reproduction of partial charge data or electrostatic potentials is affected by these vdW parameters.

$$E_{vdW} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$$

Eq. 4

Several other parameters should also be considered, such as the out-of-plane bending and cross-term parameters. The out-of-plane bending can be handled in a variety of ways, but the use of an improper torsion is common. Regardless of the functional form, these parameters affect the reproducibility of Hessian elements. Stretch-bend, bend-bend, the various torsion cross terms and other common cross terms are often avoided to achieve the simplest and most general force field.

Defining the force field

One of the significant differences between electronic structure theory and molecular mechanics is the need for bond and atom types, which describe the connectivity of a molecule. In the context of common implementations of MM methods, differentiation of parameters is based on this connectivity. For example, different parameters will be assigned for a $C_{sp2}-C_{sp2}$ single bond versus a double bond. Additionally, MM methods may introduce zero-order bonds, but evaluation of these interactions may differ from standard bond orders. Angle-angle cross terms are not included, and 1,4-interactions are not softened in MacroModel's implementation of zero-order bonds.

When developing a TSFF, it is recommended to use bonding similar to the reactants and to incorporate zero-order bonds for bond forming/breaking interactions. Zero-order bonds can also be used for π -bonds to metals such as in allyl complexes. Aromatic rings that are π -bound to a metal have a variety of bonding descriptions, but we favor a metal-bound dummy atom that is zero-order bound to each of the vertices of the ring.⁶⁰ This has been shown to be the most realistic description of the structure and dynamics of these complexes, while the large parameter space can be handled by the automated parameterization in Q2MM.

The addition of new differentiated interactions will necessitate addition of new parameters to the force field. It is frequently necessary to override existing general parameters from the default force field that are not relevant to the new environment in which they are placed. In MacroModel, we favor adding all new parameters relevant to one type of metal complex or transition state in a *substructure* at the end of the force field text file, where the new parameters will override any other interactions that may have been assigned. A typical example is the angle around an atom that is rehybridized in the reaction studied. The bonding used in the TS model will define one hybridization, and a general parameter for that hybridization type is assigned, even if it may not be appropriate in the new context. In general, new parameters should be assigned to all new bonds and to all angles and torsions incorporating such a bond.

Estimation of initial values for new parameters

Concurrently with adding new interactions and parameters, parameters are initialized to reasonable starting values. This provides an advantage in the optimization since fewer iterations should be required. As such, initial values for bond lengths and angles should be set to the average values obtained from the model systems. Force constants should initially be set to standard values, usually 5 mdyn/Å and 0.5 mdyn/rad in MM3, for all bonds and angles, respectively.

Torsional amplitudes require a different approach. Due to 1,4-nonbonded interactions, amplitudes are typically fit to the full rotational profile of the torsion. Such a profile cannot be obtained if the torsion is part of ring systems or if specific values are required by the TS geometry. Additionally, the strong linear dependence of the V_n terms (Eq. 3) introduces the potential for overparameterization. For these reasons, a single V term is used initially and only in the case when other parameters are unable to account for this interaction are additional values added. These values should be initialized at relatively low values (near zero) as to not affect initial parameter optimization. Any additional parameters (cross terms, etc.) are, in general, rarely used and thus set to defaults or zero.

Of the two types of nonbonded interactions, only the electrostatic parameters are optimized. Electrostatic parameters are always fit first and therefore do not require any special initialization procedure. However, it should be noted that dipoles in a ring system have an infinite number of solutions, and thus one of the dipoles should be removed from the parameterization (default to zero). We emphasize avoiding the fitting of vdW parameters for a variety of reasons. First, both the electrostatic and vdW terms are linearly dependent with respect to the Hessian, and therefore multiple parameter sets can be acquired. As a result, it can be difficult to find the most accurate parameter set. Second, vdW interactions have been well defined for common force fields where individual radius and energy terms have been parameterized for each atom type.⁵⁸ Furthermore, fitting vdW parameters to data acquired from DFT methods is not advised (*vide supra*).⁴⁵

In addition to refining parameters of new interactions, some parameters that already exist in the force field may also need refinement. For example, bonds that are conjugated with the reaction center may require modification and should be evaluated on a case by case basis. If any bond is included in the parameters to be refined, angles including that bond should also be considered. The values of these parameters need not be reinitialized if they already exist in the force field. Similarly, any parameter can be initialized to values of related interactions that are already known. This includes borrowing values from typical ground state parameters or previously developed TSFFs.

Penalty function

Once parameters are added to the force field, they are refined by minimizing the penalty function, χ^2 (Eq. 5). For a given data point, x_i^o is the reference value (from QM model calculations), x_i is the value calculated using the MM parameters and w_i is a weight given to this type of data. The

weight can be determined by a number of factors, such as the reliability of the data or the estimated importance to the final usage of the force field. Weights for geometric parameters are based on the acceptable error of crystal structures. Typical weights for bond lengths (± 0.01 Å), angles ($\pm 0.5^\circ$), and torsions ($\pm 1^\circ$) are thus 100 \AA^{-1} , 2°^{-1} , and 1°^{-1} , respectively. The two ways of fitting the Hessian matrix results in different ways of assigning the weight of the elements. The first method treats every matrix element the same and therefore assigns a global weight of $1 \text{ kJ/mol}^{-1} \text{ \AA}^{-2} \text{ amu}^{-1}$. In early work, the weights were sometimes modified depending on the number of separating bonds, following Hagler and coworkers.⁵⁰ Our second, updated method contains three different types of eigenvalues. The off-diagonal elements of the QM eigenmatrix should be zero, where deviations in the MM eigenmatrix represent variance in vibrational modes between the MM and QM system. These are given a relatively low weight of $0.05 \text{ kJ/mol}^{-1} \text{ \AA}^{-2} \text{ amu}^{-1}$. The diagonal elements should contain a single negative eigenvalue (reaction coordinate), which has a weight of zero, allowing for a “natural fit” of relevant parameters.⁵⁴ All other diagonal elements represent the other normal modes and are assigned a weight of $0.1 \text{ kJ/mol}^{-1} \text{ \AA}^{-2} \text{ amu}^{-1}$. A good fit of the diagonal elements is what allows a Q2MM force field to represent the energy cost of distortions perpendicular to the reaction coordinate with high accuracy, an essential element for accurate predictions.

Following this principle, a fully parameterized TSFF should ideally return a penalty value less than N , where N is the total number of data points. The weights can also be adjusted based on the number of data points of a given type. An example would be the very large number of Hessian data points used in determining force constants; where the weight is lowered to avoid overwhelming other types of data in the refinement. As a guideline, the total contributions from different types of data to the overall penalty function should be similar.

In addition to variance in reference data reproduction, the penalty function can be used to tether parameters. This is often used in the case of optimizing equilibrium angle parameters to achieve the lowest strain parameter set. Parameter tethering is the difference between the parameter value and a tethering value (e.g. observed average angle) using a low weight.

The penalty function is used to measure convergence of the automated numerical optimizations implemented in Q2MM. Parameter refinement will continue until the change in penalty function is negligible. Optimization methods include simplex, SVD, and other gradient based methods. Once the force field has been optimized, it can then undergo external validation where it is compared to experimental data.

In a simplified picture, if the response of data to parameters is monotonous and the parameters are independent, the penalty function should have only one optimum. This is obviously not the case; parameters can become dependent (*vide infra*). To address the problem of multiple minima of the penalty function, we follow a specific order of parameterization, and inspect the parameters at each

step to identify physically unreasonable values, in particular for angle parameters.

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

Eq. 5

Parameterization order

With all parameter types initialized, the optimization can then be carried out (

Figure 6). It is advisable to work with one parameter type at a time and closely monitor the results at the initial stages of parameter refinement. Since parameters are interrelated, the order in which the parametrization is performed may influence the quality of the force field produced. Additionally, breaking the procedure into two cycles can be beneficial.

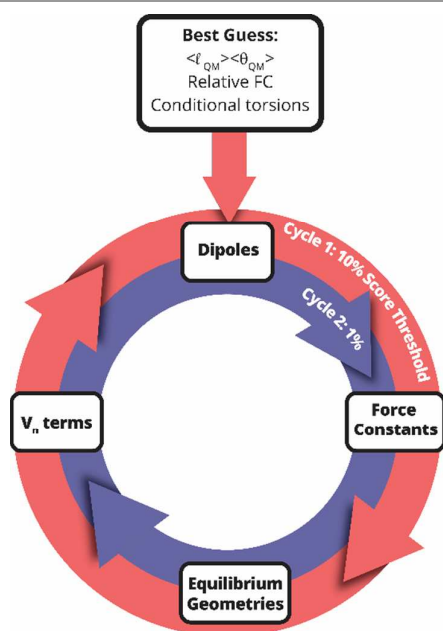


Figure 6. Flow chart of the steps taken in the Q2MM method

Cycle 1. As mentioned above, the first parameters that should be optimized are the electrostatic parameters. These parameters are mainly fit to QM partial charges, but a separate refinement round including Hessian data could be useful. Next, force constants for bonds and angles should be refined to the Hessian exclusively. Care should be taken to ensure that these force constants do not achieve very low values. Bond force constants below 1.0 mdyn/Å should be addressed, and they should never reach a value lower than 0.1 mdyn/Å. Low angle force constants can often be reasonable, but should never be lower than 0.1 mdyn/rad to avoid undue distortions in optimized structures. The equilibrium bond lengths and angles are optimized after the force constants. These parameters should be broken down into separate optimizations where they are fit to the reference bond lengths and angles, respectively. At this stage, it is crucial to pay attention to the values produced by the optimization.

Problematic parameters are often tethered to avoid exploring irrelevant parameter space (e.g. equilibrium angles should be tethered to avoid problems discussed earlier). After bond lengths and angles have been refined, users working with MacroModel are advised to run a second optimization of the electrostatic parameters, since these are represented by dipoles, which are influenced by the length of bonds that may have been changed since the initial optimization. The next stage is the refinement of proper and improper (out-of-plane) torsions. The V_n terms can be optimized similarly to the electrostatic parameters, where initial refinement involves fitting to the reference torsion values followed by fitting to the Hessian. Finally, all other correction parameters (e.g. cross-terms) should be fit to Hessian data. Relative energies can be included in parameter refinement as needed, however this may result in overfitting of parameters.

Cycle 2. The second cycle of optimization is used to further refine parameters and a possibility to add/remove some terms to improve the force field. The order of parameter types can be the same as in the first cycle, but the convergence criteria of the optimization should be decreased. By default, Q2MM will continue optimizing a parameter until the current score is within 10% of the score in the previous step of the procedure. However, in the second cycle, this threshold should be reduced to 1% to obtain better convergence. At this stage, problems with ill-defined parameters should be resolved. For example, a generic atom type could be split into two more specific ones. Improper torsional parameters and cross-terms can be considered as correction terms that should only be added to the parameterization after the core terms such as bond lengths and angles have been reasonably well converged.

Additional optimization cycles should be performed if necessary. Once a working version of the force field has been obtained, it is good practice to perform an analysis of parameters to identify deficiencies in parameter selection and topology. For example, the sensitivity of each parameter can be probed by arbitrarily increasing the convergence threshold in the optimization and verifying that the resulting values are similar to the initial ones. A global optimization could also be done but in practice seldom leads to better parameters.

Analyzing the training set data

After each parameterization refinement round, it is essential to analyze how well the objective of that particular round has been fulfilled and to understand the remaining deviations. As already stated, the penalty function should be smaller than N . This can frequently be achieved for ground state force fields, but in our experience, this level of accuracy will rarely be reached for a TSFF. Within Q2MM, each large remaining error is analyzed to evaluate whether it will have a serious influence on the final intended use of the force field. In this section, we will give examples of this type of judgement, based on data types, and how to address deviations that are judged to be unacceptable.

Parameter ranges. As part of the Newton-Raphson optimization, the sensitivity of the penalty function with respect to each parameter is determined. It is possible to

utilize these data in a back-transformation to estimate how well determined each parameter is at the end of a refinement, where each parameter should be in a penalty function minimum. Arbitrary maximum allowed increases of the penalty function (typically 1% or less) are set, followed by calculation of the variation of the parameter that would result in this increase.⁶¹ Evaluating this range of the parameter can, in some instances, determine if the inclusion of the parameter in the force field is justified. For example, if the range of a torsional parameter brackets the value zero, the use of the parameter is not justified. It could be set to zero and excluded from further refinement. Another example could be when separate parameters for bonds of the same type in different environments are used, but the ranges of the two parameter sets overlap. In this case, it could be easier to simplify the force field by merging the two parameters.

Bonds. A general deviation that is bound to occur with any TSFF is that bond lengths connected to the reaction coordinate will vary strongly in different model QM calculations, and this variation cannot be reproduced by a TSFF.⁴¹ Thus, such bonds will always show strong deviations even in final force fields. The best that can be done within the Q2MM method is to apply a high force constant to ensure all such bonds come out at approximately the same value.⁵⁴ For most other bonds, a close reproduction should be achievable. Alternatively, bonds have to be environment-dependent, and different parameter sets are used depending on the surrounding structure. This type of differentiation is easily realized in MacroModel using the substructure mechanism.

Finally, it could be the case that the MM- and QM-calculated bond lengths of one specific type correlate well but with a slope different from 1. If the range is wide, this usually means that the same bond type is found in environments that differ strongly in imposed strain, and that the force constant needs modification. Generally, force constants should be set based on vibrational data, but in a case like this, it could be modified by changing the equilibrium value and force constant simultaneously.⁶¹

Angles. In the post-refinement analysis of angles, equilibrium angle parameters that differ strongly from the average of observed values for that type or have force constants with unreasonably low values should be reevaluated. In such cases, it is generally advisable to reset the parameters to the initial values, apply tethering, and re-run the parameterization.

This typically occurs if the bond angles couple with other angles that are set in the underlying force field but not included in the parameterization. In such cases, it can be beneficial to override the problematic parameters and determine new values for them in the new context.

Torsions. These parameters are not always easy to fit well. The available torsional parameters only allow for minima at certain angles, and the observed torsions are generally strongly influenced by nonbonded interactions. Torsional parameters should be motivated by underlying physical interactions. If parameters vary to physically implausible

values, the error should frequently be sought in the nonbonded interactions instead.

Hessian data. As opposed to observables such as bonds, where generally only lengths of the parameters for the bond that is being refined are compared, Hessian data from QM and MM calculations are compared for the entire molecule. This can be done using a transformation to QM eigenvectors,⁵⁴ which simplifies the analysis. Inspection of the normal mode vibration corresponding to each eigenvalue informs whether deviations arise from force constants included in the refinement or not. Large MM values for off-diagonal elements (which should be zero if the eigenvectors correspond perfectly between QM and MM) indicate whether cross-terms are needed. In MM3, a stretch-bend term can be included if the off-diagonal element connecting a stretching and a bending vibration gives a large deviation in the post-refinement analysis.

Comparison to experimental validation set

A TSFF force field is intended to predict relative rates for the formation of stereoisomers. We intentionally do not use any experimental data in the force field refinement stage but rather validate by calculating selectivities and comparing to experiments. Comparisons are best done on the energy scale, equivalent to the $\ln(e.r.)$ scale, where we expect errors to be similar over the entire range of selectivities (*vide infra*). Computationally, the potential energies from all calculated transition states are Boltzmann averaged for all paths leading to one isomer. This determines the ratio between isomers. From this final ratio, we calculate the energy difference, and compare to the experimental energy difference calculated similarly from the experimental ratio.

In earlier implementations, this was sometimes simplified by selecting only the lowest energy conformer leading to each isomer. The two procedures give very similar results, but the Boltzmann summation has the advantage of also including conformational entropy in the final energy difference, making it more directly comparable to the experimental free energy differences.

Q2MM Standard usage

We have recently reviewed the standard usage of Q2MM,⁹ so only a brief summary will be given here. Applications beyond standard usage are covered in a separate section. Three of the reactions we have studied fall under this standard usage heading:

- Osmium-catalyzed asymmetric dihydroxylation.⁶²
- Rhodium-catalyzed asymmetric hydrogenation of enamides.⁶³
- Ruthenium-catalyzed hydrogenation of ketones.⁶⁴

Our main applications have focused on selectivity, not reaction rate, and the stereoselecting step studied is not necessarily the rate limiting one. In the presence of a chiral, enantiopure catalyst (generally derived from combining a transition metal with a chiral ligand), the transition states leading to enantiomeric products will be diastereomeric, and

the relative energies of the transition states will determine the ratio of products. In all three cases, the mechanism had already been characterized by QM methods, and the stereodetermining step had been identified. We have traditionally derived parameters from small model systems calculated using a standard DFT functional such as B3LYP⁶⁵ with a double- ζ quality basis set, *in vacuo*, and compared to MM calculations for identical systems run under the same conditions (*vide supra*). The type of QM method used to locate the selectivity-determining transition state could vary with the type of mechanism study, but a discussion of the pros and cons of different functionals for different mechanism types is beyond the scope of this review. We note that the ability of a DFT-derived Q2MM force field to reproduce experimental data in itself is a validation that the mechanism of the underlying DFT study is correct. We have used this argument to validate the mechanism of the osmium-catalyzed dihydroxylation,⁶² as well as to choose the selectivity-determining step for the Rh-catalyzed enamide hydrogenation.⁶³

For each force field, internal validation verifies that we have introduced sufficient parameters to reproduce key QM data. An external test set is gathered from published experimental examples, preferably including both low and high selectivities. Finding a suitable set of experimental data can sometimes be challenging because of a bias towards publishing the most selective results without including lower selectivities. For all entries in the test set, a prediction is calculated by performing a full conformational search on each substrate-ligand combination, usually via a combined Monte Carlo/low mode conformational search. This is a key step; failures in reproducing experimental data at early stages of development can frequently be traced to insufficient conformational searches that have been unable to locate all contributing paths.

At the end, a Boltzmann summation over all conformers leading to each stereoisomer of the product yields the final predicted difference in energy of the transition state ensembles, from which enantiomeric excess can be derived. We can also obtain a final energy difference from the expression $\Delta E = RT \ln(e.r.)$. The Boltzmann summation is based on potential energies from the force field, ignoring any contributions from differences in vibrational entropies. However, the final energy (ΔE) has some character of free energy, since a conformational entropy term is accounted for by the summation.

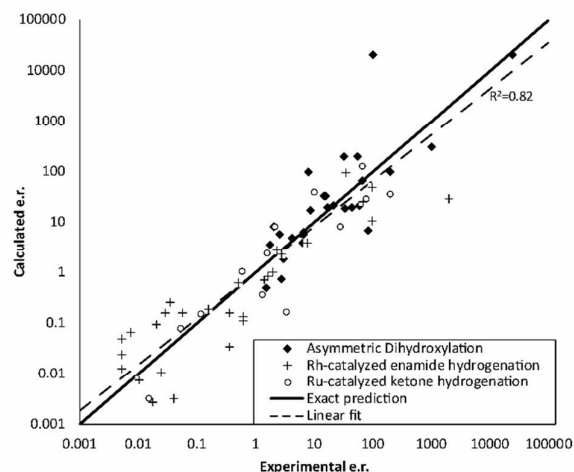


Figure 7 Predicted vs. experimental *e.r.* for our three standard use reactions.

The final results, based on the three reactions considered here, can be seen in Figure 7. The solid diagonal is not a fitted line, but a representation of perfect correspondence. In energy terms, the average error is 2.8 kJ/mol ($R^2=0.82$), with errors evenly distributed over the range. Note that due to the exponential relationship between energy and product ratio, a constant error term in energy will correspond to a constant factor of *ca.* 3 for *e.r.* This is accurate enough to be useful in an experimental context; a system that has been predicted to have an *e.e.* of 99% will on average be observed in the range 97%–99.7% in the lab, whereas a prediction of 0% *e.e.* would be expected to have an experimental *e.e.* <50%. If the experimentalist's primary goal is to identify a top-performing catalyst, there would be little reason to experimentally test the cases for which low selectivity is predicted where the numerical error in *e.e.* might seem large.

Scope of a Q2MM force field

A Q2MM force field is valid for one particular reaction. New parameters have been defined for a reaction core consisting of all atoms and bonds that are unknown in the underlying ground state force fields. For example, the Rh-catalyzed hydrogenation is defined for a reaction core including the enamide, and will not be valid for other types of alkenes. The Ru-catalyzed hydrogenation is more general with respect to the substrate, but limited to catalysts with two coordinating phosphines and two coordinating amines. Changing the nature of atoms or bonds in the core will require deriving a new force field.

Each force field is derived for a selected set of substrates (and ligands). Other substrates can easily be used directly as long as they can be calculated by the underlying ground state force fields but changes close to the reaction center may require optimization of a small number of new parameters. For example, the dihydroxylation force field was derived for tertiary amine ligands on Os. To allow pyridine ligands, parameters for an sp^2 -nitrogen bound to the metal, as well as angles and torsions including this bond, would need to be added to the force field, but the reaction core including

the alkene and osmium tetroxide could probably be used without modification. In the same reaction, parameters were derived for alkyl and aryl substituents on the reacting alkene, but even a simple benzyl substituent uses a few low quality, generic angle parameters from the underlying MM3* force field. Improving the results for such substrates would require refinement of a few parameters.

Challenges to Q2MM

Like all force field methods, Q2MM comes with a number of inherent limitations in scope. Most importantly, force field comparisons are only valid for comparing two structures that have the exact same types of atoms and bonds. In a general sense, these comparisons are called *isodesmic*, but when applied to force fields, we use the term *isoparametric*: the same number of parameters of each type must be used for the two systems being compared. This generally limits the usage to comparisons of conformers or stereoisomers. Additionally, when transition states are approximate as an energy minimum, results can deteriorate if the transition state structure is not relatively rigid and constant. Shifts of the transition state along the reaction coordinate, for example when subjected to extreme steric hindrance, are not captured by an energy minimum model.^{40,41} However, in the two decades since the invention of the Q2MM method,⁸ many of these limitations have served as challenges to overcome. In this section, we will summarize some of the lessons learned from these attempts.

Loose transition states (i): dialkylzinc addition to aldehydes

One of the most intensely studied of the early asymmetric reactions was addition of diethylzinc to benzaldehyde, catalyzed by chiral amino alcohols. The mechanism, as revealed in a seminal computational study by Yamakawa and Noyori,⁶⁶ features a bifunctional alkyl-zinc-alcoholate complex bringing together an aldehyde substrate with a dialkylzinc reagent in a tricyclic transition state (Figure 8). The original study, as well as later contributions,⁶⁷ revealed several plausible transition state geometries, which could become accessible as reaction components change. A Q2MM force field was derived based on model calculations for 17 tricyclic transition states.⁶⁸ Due to the variable nature of coordination around the two metal atoms, fitting of force field parameters was unusually challenging. Extensive harmonic tethering, especially of angular parameters, was required to derive a model that could reproduce the training set acceptably while retaining physically sensible values for all parameters.

The final force field was tested on ten diverse combinations of four aldehyde substrates and seven amino alcohol ligands. Gratifyingly, the final force field could identify the preferred enantiomer in all cases, but the numerical accuracy was lower than we have achieved for mono-metallic catalysts (Figure 7). The mean unsigned error over all test examples was 4.3 kJ/mol. The accuracy was good enough to allow an identification of which of three possible pathways can contribute to the minor enantiomer for each ligand,

information that is important in the design of improved ligand systems.

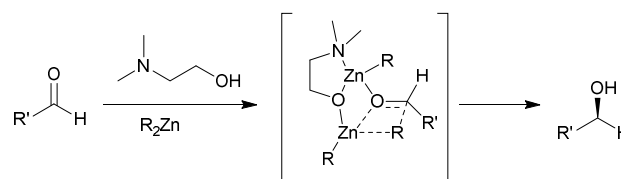


Figure 8 Dialkyl zinc addition to aldehydes.

Loose transition states (ii): Mukaiyama aldol reaction

It is generally assumed that high stereoselectivity requires multiple close contacts between the substrate(s) and the chiral inducer. However, the Mukaiyama aldol reaction of an aldehyde, a silyl enol ether, and a boron catalyst proceeds through an open TS with only one strong contact between the two substrates (the forming C-C bond) and one contact with the chiral inducer (an O-B coordination of the aldehyde electrophile to the catalyst).¹⁵ In addition, due to the open nature of the TS, the reaction undergoes a strong charge shift. Three neutral components come together and generate a complex that consists of a negative boronate at one end and a positive silyl cation at the other end (Figure 9). In a later stage, the silyl group is transferred, liberating a neutral boron catalyst and a neutral silyl ether product. Modeling a reaction coordinate that involves a separation of charge is challenging for all methods.

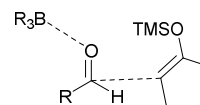


Figure 9 Mukaiyama aldol reaction

A Q2MM model was developed and applied to a model reaction. From extensive conformational searching, it was possible to identify a non-classical hydrogen bond that contributed secondary interactions, stabilizing one isomer preferentially. However, due to the aforementioned challenges, it was not possible to achieve a quantitatively correct model given that calculated selectivities differed from the experimental values by ca. 6-8 kJ/mol. Using the Q2MM force field as a conformational search tool, starting structures for DFT optimizations were generated because single point DFT calculations using Q2MM did not offer any improvement. These structures were excellent starting points for TS optimizations, allowing for a reliable and fast location of a number of transition state conformations using DFT. The DFT structures confirmed the secondary interactions identified by Q2MM and gave a close agreement for the diastereoselectivity of the reaction. Disappointingly, the combination of methods did not allow a correct identification of the preferred path to the minor enantiomer because the enantioselectivity was overestimated by >10 kJ/mol. We hypothesize that the best TS for the minor enantiomer may exist in the Q2MM conformational search, but due to the inaccuracies of Q2MM

for this challenging TS, it is above the energy cutoff for selecting starting structures for DFT optimization.

Multiple selectivity-determining steps: the HWE reaction

One of the basic tenets of force field calculations is that comparisons of structures with different bonding gives undefined results, since the difference in bond dissociation energy is ignored by most force fields. A corollary in Q2MM should be that reactions for which the results are influenced by multiple transition states should be out of scope. However, the first Q2MM application addressed a reaction with two potentially selectivity-determining steps: the asymmetric Horner-Wadsworth-Emmons reaction.⁵¹ In this reaction, an enantiopure phosphonate reagent is added to a racemic aldehyde, creating an intermediate with two additional stereogenic centers. The subsequent elimination is diastereospecific, eliminating the newly created stereocenters but yields an alkene where the *E/Z* selectivity is determined by the relative stereochemistry in the preceding intermediate. Both transition states had been found to have similar energies,⁶⁹ and therefore it was not *a priori* clear which of them would be selectivity determining (Figure 10).

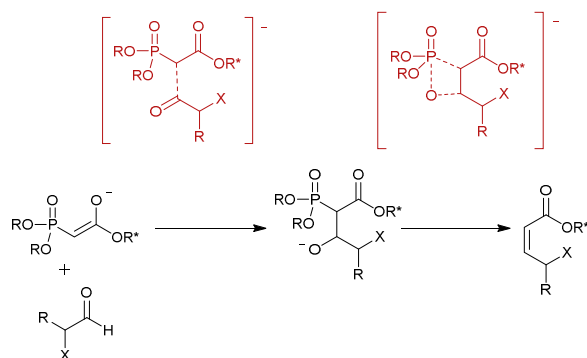


Figure 10 Asymmetric HWE reaction with two selectivity-determining transition states

Force fields were created for both transition states with the initial aim to test the selectivity in each TS and to compare computational and experimental results. It became clear that both transition states exert an influence on the final selectivity. With the additional stereocenters available in the intermediate, the reaction proceeds through eight diastereomerically distinct paths, each with multiple contributing conformations. If either of the two subsequent transition states on a path has a high energy, the addition will be blocked and revert to starting materials. The energy ranking within each type of transition state could be determined accurately from the force fields, but the relative energy of the two classes of transition states had to be obtained from other sources.

To further complicate matters, the unknown correction for the relative energies of the two types of transition states must be assumed to be distinct for each substrate. One option was to determine a correction factor by taking one conformation from each type of TS and determine the actual relative energy by high level QM calculations. However, at the time of this

work (before the advent of dispersion corrections), QM methods were neither fast nor accurate enough for this approach to be practical. Instead, the unknown energy correction was used as a fitting parameter, varied to achieve the best possible correspondence with experimental selectivities.

For this approach to be successful, there must exist a combination of the two classes of transition states that matches experiments. Furthermore, each experiment yields three distinct selectivities (*E/Z* ratio and enantiomeric ratio for each alkene isomer), and all three selectivities must arise from the same model using only one energy correction. The final model yielded a mean unsigned error of 1.6 kJ/mol. Gratifyingly, the energy corrections were found to vary with the bulk of the reagent, increasing the energy of the more compact elimination TS more for the bulkier systems, as expected.

We note that when used in this mode with experimental data as anchor points in the evaluation, the model is no longer truly predictive. As with most computational studies, the main use is to reveal the factors influencing selectivity and to generate an in-depth understanding that will aid future development. This will be sufficient for many applications, but for a truly predictive method, it is necessary to find an independent means to relate the two types of transition states to each other. The most obvious would be to use the Q2MM force fields as search tools to generate the best conformations for each TS and then to compare them to each other at an appropriate QM level. It should be stressed that each TS must be reoptimized with the chosen QM method in the manner employed for the Mukaiyama reaction (*vide supra*). Due to the known inability of Q2MM force fields to reproduce distortions along the reaction coordinate,⁴¹ (*vide supra*) single point calculations using Q2MM geometries are not reliable for generation of accurate energies.⁶⁷

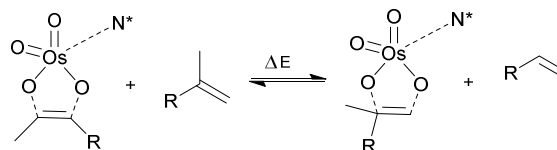


Figure 11 Isodesmic comparison of two dihydroxylation reactions

Addressing relative reactivity: isodesmic dihydroxylation reactions

Computational predictions in general become more accurate when the chemical difference between two systems being compared is minimized. In particular, isodesmic comparisons can be done successfully using relatively low levels of theory.⁴³ An isodesmic comparison is generally implemented as an exchange reaction (not necessarily chemically realistic) where exactly the same bond and atom types are present on both sides of the equilibrium. For chemical reaction rates, isodesmic comparisons are generally based on a pseudo-equilibrium where a transition state for one substrate plus a ground state for a second substrate is compared to a state where the two substrates have changed

position (Figure 11). For comparisons using force fields, the requirements are more stringent; the comparison must be *isoparametric*. The same number of all parameters must be used on both sides of the equilibrium.

The concept of isodesmic comparisons for relative rates of reactions has been tested using Q2MM methods for the asymmetric dihydroxylation reaction (AD). Using a force field that was originally developed to reproduce the stereoselectivity of the reaction⁶² (Figure 7), the relative dihydroxylation rates of isomeric trisubstituted alkenes were predicted.⁷⁰ All comparisons were isodesmic but not fully isoparametric; there were minor differences in which angle and torsional parameters were utilized by the different substrates. For example, for three different dimethyl styrenes, two substrates would have a Me-C(sp²)-Ph angle, whereas the remaining substrate would have a Me-C(sp²)-Me angle. Despite these differences, the predictions were highly successful, reproducing the relative rates of the different substrates with an accuracy similar to what was obtained for the stereoselectivity predictions. It could be shown that the relative rates depended mainly on the ability of the ligand to stabilize the substrate by nonbonded interactions, in line with earlier kinetic studies.⁴⁶ The results also allowed an update on the selectivity mnemonic for the AD reaction with a full rationalization of which parts of the ligand caused each interaction represented by the mnemonic.

Reactions in enzymes: P450 reactivity

The most common application of force fields has historically been biomolecules.⁷¹⁻⁷⁴ Their sheer sizes make them impractical to treat the entire system with quantum mechanics, especially on the timescales relevant for addressing biological questions. The force fields that are used for proteins are all used to simulate the ground state and by their nature cannot be used to analyze chemical events. Traditionally, this problem has been overcome by the use of QM/MM models, which model the active site using QM methods, and the rest of the enzyme using MM.^{75,76} The implicit assumption is that the ground state structure of the enzyme is essentially the same conformation that the protein adopts in the transition state, which is not necessarily the case. Even in the ground state of some proteins, small changes in chemical bonds can result in significant changes in conformational preferences.⁷⁷

Q2MM has been expanded to develop a TSFF for the cytochrome P450 (CYP) family of enzymes, which are heme-containing proteins participating in many biological oxidations proceeding via electron transfer pathways.⁷⁸ Oxidation by P450 is a complex mechanism, but as long as the selectivity determining step can be ascribed to a single type of TS on one spin surface, it is amenable to TS theory and to modeling by Q2MM force fields. The derived TSFF enabled docking of multiple substrates in multiple conformations. Traditional docking procedures have used ground-state molecules and reaction centers, whereas a Q2MM TSFF selectively looks for substrates that bind to the oxoferryl group in the transition state for this system. This provides a more accurate

assessment of the reaction center and aids in a better chemical understanding of the active site. In humans, there are almost 60 genes for CYPs, but the superfamily of monooxygenases have a highly conserved active site,⁷⁹ which is ideal for parameterization (Figure 12). CYPs have also been found to be responsible for 75% of Phase I drug metabolism, making them important topics in pharmaceutical research.^{80,81}

The initial parameterization was based on the DFT optimized structures of 24 organic substrates with 14 compounds in the training set (Figure 12), and 10 compounds in the validation set. The force field on which the parameters were based was the General Amber Force Field (GAFF) in conjunction with the antechamber program in the Amber suite to ensure that atom types were as general as possible.⁸²

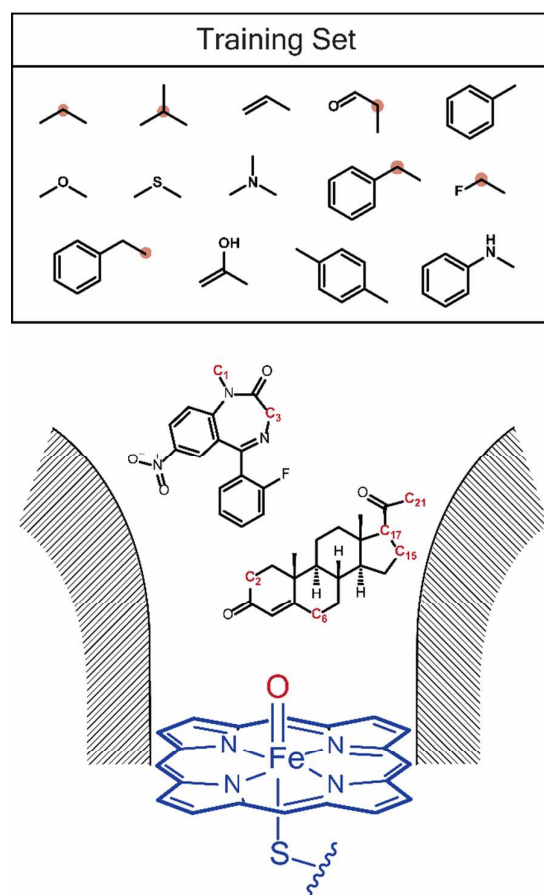


Figure 12 Top: Small molecule organic training set for a general CYP active site TSFF. Specific reacting carbons highlighted in red. Bottom: Substrates studied with the considered carbon atoms indicated.

The minimization of the penalty function is similar to that of small molecules. The deviation of geometries and Hessian elements between the DFT and MM data is nearly identical, but the structural data and start files for MM minimization needed to be altered due to the differences in the Amber FF and program. The most significant change stemmed from the need to modify the *nmode* module in the Amber program in order to write the forces and the mass-weighted Hessian matrix to a designated file.

The TSFF could reproduce the geometries of the validation set to a high accuracy, especially around the heme ring given the similarity between the structures. The largest discrepancies occurred for the dihedrals to the substrate and to the cysteine ligand. While the geometries were similar, the MM optimized structures gave poor energetic results when they were not re-optimized using DFT. The substrates were the largest source of the error even though they were not reparameterized from their GAFF values, and caused a discrepancy of about 7-22 kJ/mol, with the overall error around 14-37 kJ/mol compared to DFT.

Despite this error in energy, the Q2MM TSFF can be used to determine if a substrate can sterically fit into the active site and could be a substrate of the enzyme. To test this force field for docking specific drugs, progesterone and flunitrazepam were studied⁸³ (Figure 12).

The drugs were manually docked into the active site in a reasonable conformation with the reactive hydrogen atom close to the oxoferryl group. This complex was then minimized keeping the protein and heme group fixed and allowing the substrate to move. This was followed with a full conformational analysis of the substrates by rotating the four rotatable bonds (Fe-O, O-H, H-C, and C17-C20). These structures were optimized using molecular mechanics with the protein and heme group fixed. The lowest energy structure from this search was used for analysis for binding and estimation of the binding affinities.

The conformational sampling gave qualitatively sufficient results and provides better predictions for sites of metabolism by CYPs than the scoring with GOLD program and metabolic predictions by MetaSite.^{84, 85} The errors in the energy mentioned above often translates into incorrect ranking of sites for different CYPs. Given that the Q2MM force field energies cannot be compared to the ground state force fields used in other software, the results were compared by the number of misses for the various sp³ carbons. In all cases but one, the Q2MM force field correctly predicted the misses, and was significantly better than the common docking software in use at the time.

Outlook

Improved methods

Over the years, the main focus of Q2MM development has been on the principles of developing force fields. The parameterization package has been advanced to the “good enough” level, but could be further improved. At the current stage, the logical next step is to make the Q2MM package more efficient while still delivering the same quality of force fields. These steps include replacing some of the numerical differentiation steps with analytical expressions, implementing a conjugate gradient optimizer, and developing a package for automated initialization of new force fields, including initial parameter estimation.

Addressing the reaction coordinate response

The major advantages of the TSFF method when derived using Q2MM are the ability to search for reaction paths using reliable conformational search tools and the accuracy of the energy response to distortions perpendicular to the reaction coordinate. However, displacements along the reaction coordinate are not correctly represented.⁸ Other techniques, such as EVB and SEAM, will allow a proper reaction coordinate response. An obvious extension is therefore that the Q2MM structures could be refined using either of these techniques. An option that may be even better is to employ ACE, which could conceivably be reparameterized at the TS to give a performance similar to Q2MM, and also allow explorations along the reaction coordinate.

QSAR reactivity models based on Q2MM

QSAR-based reactivity models (QSRR) have reemerged as a very promising technique in supporting reaction development.¹² However, most such models are based on data derived at the ground state. We have previously shown that a force field model of transition states (a predecessor of Q2MM) can deliver data that is useful in building QSRR models.⁸⁶ We envision that full Q2MM models could deliver information about the transition state useful in QSRR or machine learning models, enabling us to go beyond the force field limitation to isomers only and starting to address reactivity.

Conclusions

Over the last two decades, Q2MM force fields have been used to extrapolate from QM calculations to prediction of experimental stereoselectivities with an accuracy that is at this point unmatched by any other approach. The major factors contributing to the success are:

- Basing the transition state force fields on an accurate underlying ground state force field working from physically sound principles.
- Closely reproducing DFT transition states for model systems that are small enough not to suffer from errors due to lack of dispersion or solvation.
- Representing the transition state as an energy minimum, allowing a complete determination of all contributing pathways through efficient conformational searching.

The Q2MM package providing the functionalities discussed in this contribution are, together with the validated force fields derived by Q2MM, available to the scientific community free of charge at github.com/q2mm. Use of these TSFFs to accurately predict the stereoselectivity of a wide range of reactions and the ability to quickly generate new force fields for reactions of interest using Q2MM provides an important new tool for the organic chemistry community that is an important step towards the “Holy Grail” of predictive computational modeling in catalysis.⁸⁷

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

We gratefully acknowledge support of the Q2MM project by the National Science Foundation (CHE-1565669 and OISE-1261033), the National Institutes of Health (1R01GM111645) and Shenzhen STIC (JCYJ20170412150343516). A.R.R. and T.R.Q. are recipients of CBBI fellowships (NIH T32 GM075762)

Notes and references

- 1 a) *Chirality in Drug Research*. Francotte, E.; Lindner, W., Eds. Wiley-VCH: Weinheim, 2007. b) *Asymmetric Synthesis: The Essentials*. 2nd ed. Christmann, M.; Bräse, S., Eds. Wiley-VCH: Weinheim, 2008.
- 2 a) Lovering, F.; Bikker, J.; Humblet, C. "Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success", *J. Med. Chem.* 2009, 52, 6752–6756 b) Lovering, F. "Escape from Flatland 2: complexity and promiscuity", *Med. Chem. Commun.* 2013, 4, 515–519.
- 3 a) Lindsley, C.W. "New 2016 Data and Statistics for Global Pharmaceutical Products and Projections through 2017", *ACS Chem. Neurosci.* 2017, 8, 1635–1636 b) Philippidis, A. "The Top 15 Best-Selling Drugs of 2017". *Genet. Eng. Biotech. News*. March 6, 2017. <https://www.genengnews.com/the-lists/the-top-15-best-selling-drugs-of-2016/77900868>.
- 4 Federsel, H.-J.; Larsson, M. "An Innovative Asymmetric Sulfide Oxidation: The process Development History behind the New Antiulcer Agent Esomeprazole" in *Asymmetric Catalysis on Industrial Scale: Challenges, Approaches and Solutions*. Blaser, H.-U.; Schmidt, E., Eds. Wiley-VCH: Weinheim, 2004, pp 413–436.
- 5 Corey, E. J.; Kürti, L. *Enantioselective Chemical Synthesis: Methods, Logic and Practice*. Direct Book Publishing: Dallas, 2010. ISBN 978-0-615-39515-9.
- 6 *Comprehensive Enantioselective Organocatalysis*. Dalko, P. I. Ed.; Wiley-VCH: Weinheim, 2013.
- 7 S. W. Krska, D. A. DiRocco, S. D. Dreher, and M. Shevlin, "The Evolution of Chemical High-Throughput Experimentation to Address Challenging Problems in Pharmaceutical Synthesis", *Acc. Chem. Res.* 2017, 50, 2976–2985.
- 8 Norrby, P.-O. "Molecular Mechanics as a Predictive Tool in Asymmetric Catalysis" in *Transition State Modeling for Catalysis*. Truhlar, D.G.; K. Morokuma, K. Eds. ACS Symposium Series 721, American Chemical Society: Washington, DC, 1999, pp 163–172.
- 9 Hansen, E.; Rosales, A.R.; Tutkowski, B.; Norrby, P.-O.; Wiest, O. "Prediction of stereochemistry using Q2MM", *Acc. Chem. Res.* 2016, 49, 996–1005.
- 10 a) Ianni, J.C.; Annamalai, V.; Phuan, P.W.; Panda, M.; Kozlowski, M.C. "A priori theoretical prediction of selectivity in asymmetric catalysis: Design of chiral catalysts by using quantum molecular interaction fields", *Angew. Chem Intl. Ed.* 2006, 45, 5502–5505 b) Kozlowski, M.C.; Dixon, S.L.; Panda, M.; Lauri, G. "Quantum Mechanical Models Correlating Structure with Selectivity: Predicting the Enantioselectivity of Beta-Amino Alcohol Catalysts in Aldehyde Alkylation", *J. Am. Chem. Soc.* 2003, 125, 6614–6615 c) Kozlowski, M.C.; Panda, M. "Computer-aided design of chiral ligands Part I. Database search methods to identify chiral ligand types for asymmetric reactions", *J. Mol. Graph. Mod.* 2002, 20, 399–409 d) Kozlowski, M.C.; Panda, M. "Computer-Aided Design of Chiral Ligands. Part 2. Functionality Mapping as a Method to Identify Stereocontrol Elements for Asymmetric Reactions", *J. Org. Chem.* 2003, 68, 2061–2076 e) Kozlowski, M.C.; Waters, S.P.; Skudlarek, J.W.; Evans, C.A. "Computer-Aided Design of Chiral Ligands. Part III. A Novel Ligand for Asymmetric Allylation Designed Using Computational Techniques" *Org. Lett.* 2002, 4, 4391–4393 f) Lipkowitz, K.B.; Kozlowski, M.C. "Understanding stereoselection in catalysis via computer: New tools for asymmetric synthesis", *Synlett*, 2003, 1547–1565.
- 11 Cross, S.; Cruciani, G. "Molecular fields in drug discovery: getting old or reaching maturity?", *Drug Disc. Today* 2010, 15, 23–32.
- 12 a) Harper, K.C.; Sigman, M.S. "Predicting and optimizing asymmetric catalyst performance using the principles of experimental design and steric parameters", *Proc. Nat. Acad. Sci.* 2011, 108, 2179–2183 b) Harper, K.C.; Sigman, M.S. "Using physical organic parameters to correlate asymmetric catalyst performance", *J. Org. Chem.* 2013, 78, 2813–2818 c) Sigman, M.S.; Harper, K.C.; Bess, E.N.; Milo, A. "The development of multidimensional analysis tools for asymmetric catalysis and beyond", *Acc. Chem. Res.* 2016, 49, 1292–1301.
- 13 Tropsha, A. "Best practices for QSAR model development, validation, and exploitation", *Mol. Inf.* 2010, 29, 476–488.
- 14 For some recent examples from our group, see e.g. a) Xu, L.; Zhang, X.; McCammant, M. S.; Sigman, M. S.; Wu, Y.-D.; Wiest, O. "Mechanism and Selectivity in the Pd-Catalyzed Difunctionalization of Isoprene", *J. Org. Chem.* 2016, 81, 7604–7611 b) Tutkowski, B. Meggers, E.; Wiest, O. "Understanding Rate Acceleration and Stereoselection of an Asymmetric Giese Reaction Mediated by a Chiral Rhodium Catalyst", *J. Am. Chem. Soc.* 2017, 139, 8062–8065 c) Hilton, M. J.; Xu, L.P.; Norrby, P.O.; Wu, Y.D.; Wiest, O.; Sigman, M.S. "Investigating the Nature of Palladium Chain-Walking in the Enantioselective Redox-Relay Heck Reaction of Alkenyl Alcohols", *J. Org. Chem.* 2014, 79, 11841–11850.
- 15 Lee, J.M.; Zhang, X.; Norrby, P.O.; Helquist, P.; Wiest, O. "Stereoselectivity in (Acyl)borane-Catalyzed Mukaiyama Aldol Reactions", *J. Org. Chem.* 2016, 81, 5314–5321.
- 16 Chung, L.W.; Sameera, W.M.C.; Ramozzi, R.; Page, A.J.; Hatanaka, M.; Petrova, G.P.; Harris, T.V.; Li, X.; Ke, Z.; Liu, F.; Liu, F.; Li, H.B.; Ding, I.N.; Morokuma, K. "The ONIOM method and its applications", *Chem Rev.* 2015, 115, 5678–5796.
- 17 a) Reetz, M.T.; Meiswinkel, A.; Mehler, G.; Angermund, K.; Graf, M.; Thiel, W.; Mynott, R.; Blackmond, D.G. "Why are BINOL-based monophosphites such efficient ligands in Rh-catalyzed asymmetric olefin hydrogenation?", *J. Am. Chem. Soc.* 2005, 127, 10305–10313. b) Feldgus, S.; Landis, C.R. "Large-scale computational modeling of [Rh (DuPHOS)](+) catalyzed hydrogenation of prochiral enamides: reaction pathways and the origin of enantioselection.", *J. Am. Chem. Soc.* 2000, 122, 12714–12727.
- 18 Rooks, B.J.; Haas, M.R.; Sepúlveda, D.; Lu, T.; Wheeler, S.E. "Prospects for the Computational Design of Bipyridine, N,N'-Dioxide Catalysts for Asymmetric Propargylation Reactions", *ACS Catalysis* 2015, 5, 272–280.
- 19 Jensen, F.; Norrby, P.-O. "Transition states from empirical force fields", *Theor. Chem. Acc* 2003, 109, 1–7.
- 20 Warshel, A.; Weiss, R.M. "An empirical valence bond approach for comparing reactions in solution and in enzymes", *J. Am. Chem. Soc.* 1980, 102, 6218–6226.
- 21 Warshel, A.; Sussman, F.; Hwang, J.K. "Evaluation of catalytic free energies in genetically modified proteins", *J. Mol. Biol.* 1988, 201, 139–159.
- 22 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.
- 23 Hwang, J.K.; King, G.; Creighton, S.; Warshel, A. "Simulation of Free Energy Relationships and Dynamics of S_N2 Reactions

- in Aqueous Solution", *J. Am. Chem. Soc.* 1988, 110, 5297-5311.
- 24 a) Grochowski, P.; Lesyng, B.; Bala, P.; McCammon, J.A. "Density functional based parametrization of a valence bond method and its applications in quantum classical molecular dynamics simulations of enzymatic reactions", *Intl. J. Quant. Chem.* 1996, 60, 1143-1164 b) Bala, P.; Grochowski, P.; Lesyng, B.; McCammon, J.A. "Quantum-classical molecular dynamics simulations of proton transfer processes in molecular complexes and in enzymes", *J. Phys. Chem.* 1996, 100, 2535-2545.
- 25 a) Albu, T.; Corchado, J.C.; Truhlar, D.G. "Molecular mechanics for chemical reactions: A standard strategy for using multiconfiguration molecular mechanics for variational transition state theory with optimized multidimensional tunneling", *J. Phys. Chem. A* 2001, 105, 8465-8487 b) 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 c) Higashi, M.; Truhlar, D.G. "Combined electrostatically embedded multiconfiguration molecular mechanics and molecular mechanical method: Application to molecular dynamics simulation of a chemical reaction in aqueous solution with hybrid density functional theory", *J. Chem. Theory Comput.* 2008, 4, 1032-1039 d) Tishchenko, O.; Truhlar, D.G. "Efficient global representations of potential energy functions: Trajectory calculations of bimolecular gas-phase reactions by multiconfiguration molecular mechanics", *J. Chem. Phys.* 2009, 129, 024105 e) Lin, H.; Pu, J.Z.; Albu, T.V.; Truhlar, D.G. "Efficient molecular mechanics for chemical reactions: Multiconfiguration molecular mechanics using partial electronic structure Hessians", *J. Phys. Chem. A* 2004, 108, 4112-4124 f) Tishchenko, O.; Truhlar, D.G. "Global potential energy surfaces with correct permutation symmetry by multiconfiguration molecular mechanics", *J. Chem. Theory Comput.* 2007, 3, 938-948 g) Lin, H.; Zhao, Y.; Tishchenko, O.; Truhlar, D.G. "Multiconfiguration molecular mechanics based on combined quantum mechanical and molecular mechanical calculations", *J. Chem. Theory Comput.* 2006, 2, 1237-1254.
- 26 •Rappé, A.K.; Pietsch, M.A.; Wisner, D.C.; Hart, J.R.; Bormann-Rouchotte, 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.
- 27 For a discussion of the relationship between these methods, see a) Florian, J. "Comment on Molecular Mechanics for Chemical Reactions", *J. Phys. Chem. A* 2002, 106, 5046-5047 (b) Truhlar, D. G. "Reply to Comment on Molecular Mechanics for Chemical Reactions", *J. Phys. Chem. A* 2002, 106, 5048-5050.
- 28 Jensen, F. "Locating minima on seams of intersecting potential energy surfaces", *J. Am. Chem. Soc.* 1992, 114, 1596-1603.
- 29 Lomas, J. S.; Dubois, J. E. "Acid-catalyzed dehydration and acetolysis of tertiary methyl- and tert-butylcarbinols. Empirical force field treatment of tert-butyl/methyl reactivity ratios in solvolysis reactions of alcohols and p-nitrobenzoates", *J. Org. Chem.* 1979, 44, 1647-1654.
- 30 Hogg, J. L.; Rodgers, J.; Kovach, I.; Schowen, R. L. "Kinetic isotope-effect probes of transition-state structure. Vibrational analysis of model transition states for carbonyl addition", *J. Am. Chem. Soc.* 1980, 102, 79-85.
- 31 Williams, I.H. "On the representation of force fields for chemically reacting systems", *Chem. Phys. Lett.* 1982, 88, 462-466.
- 32 a) Wu, Y. D.; Houk, K. N. "Electronic and conformational effects on pi-facial stereoselectivity in nucleophilic additions to carbonyl compounds", *J. Am. Chem. Soc.* 1987, 109, 908-910 b) Ando, K.; Condroski, K. R.; Houk, K. N.; Wu, Y. D.; Ly, S. K.; Overman, L. E. "Stereoselectivities of nucleophilic additions to cycloheptanones. Experimental and theoretical studies and general purpose force field for the prediction of nucleophilic addition stereoselectivities", *J. Org. Chem.* 1998, 63, 3196-3203 c) Ando, K.; Houk, K. N.; Busch, J.; Menasse, A.; Sequin, U. "Experimental and computational studies of nucleophilic additions of metal hydrides and organometallics to hindered cyclohexanones", *J. Org. Chem.* 1998, 63, 1761-1766. d) Dorigo, A. E.; Houk, K. N. "On the Relationship between Proximity and Reactivity - An ab initio Study of the Flexibility of the .OH +CH₄ Hydrogen Abstraction Transition State and a Force Field Model for the Transition States of Intramolecular Hydrogen Abstractions", *J. Org. Chem.* 1998, 63, 1650-1664 e) Nakamura, K.; and Houk, K.N. "Substituent Effects on Torquoselectivity - Computational Investigations of the Thermal Ring Openings of Cyclobutenes Fused to Substituted Cyclohexenes", *J. Org. Chem.* 1995, 60, 686-691 f) Radkiewicz, J.L.; McAllister, M.A.; Goldstein, E.; Houk, K.N. "A theoretical investigation of phosphoramidates and sulfonamides as protease transition state isosteres", *J. Org. Chem.* 1998, 63, 1419-1428 g) Zipse, H.; Wang, L.H.; Houk, K.N. "Polyether catalysis of ester aminolysis - A computational and experimental study", *Liebigs Ann.* 1996, 1511-1522.
- 33 a) Eksterowicz, J. E.; Houk, K.N. "Transition-state modeling with empirical force fields", *Chem. Rev.* 1993, 93, 2439-2461 b) Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J.-R.; Hwang, M.-J.; Li, H.; Liang, C.; Maple, J.; Peng, Z.; Stockfisch, T. P.; Thacher, T. S.; Yan, L.; Ni, X.; Hagler, A.T. "Derivation of class II force fields. VIII. Derivation of a general quantum mechanical force field for organic compounds", *J. Comp. Chem.* 2001, 22, 1782-1800 c) Lipkowitz, K. B.; Peterson, M. A. "Molecular mechanics in organic synthesis." *Chem. Rev.* 1993, 93, 2463-2486.
- 34 a) Menger, F.M.; Sherrod, M.J. "Origin of High Predictive Capabilities in Transition-State Modeling", *J. Am. Chem. Soc.* 1990, 112, 8071-8075 b) Sherrod, M.J.; Menger, F.M. "'Transition-State Modeling' Does Not Always Model Transition States", *J. Am. Chem. Soc.* 1989, 111, 2611-2613.
- 35 Eurenium, K.P.; Houk, K.N. "Rational vs. Random Parameters in Transition State Modeling - MM2 Transition State Models for Intramolecular Hydride Transfers", *J. Am. Chem. Soc.* 1994 116, 9943-9946.
- 36 a) van Duin, A.C.T.; Dasgupta, S.; Lorant, F.; Goddard, W.A. "ReaxFF: A reactive force field for hydrocarbons", *J. Phys. Chem. A* 2001, 105, 9396-9409 b) van Duin, A.C.T.; Zybin, S. V.; Chenoweth, K.; Zhang, L.; Han, S.-P.; Strachan, A.; Goddard, W.A., III. "Reactive force fields based on quantum mechanics for applications to materials at extreme conditions", *AIP Conference Proceedings* 2006, 845, 581-584 c) Chenoweth, K.; van Duin, A.C.T.; Persson, P.; Cheng, M.-J.; Osgaard, J.; Goddard, W.A., III. "Development and Application of a ReaxFF Reactive Force Field for Oxidative Dehydrogenation on Vanadium Oxide Catalysts", *J. Phys. Chem. C* 2008, 112, 14645-14654 d) Jacobsson, M.; Osgaard, J.; Abrahamsson, C. O.; Norrby, P.-O.; Goddard, W.A.; Ellervik, U. "Acid-catalyzed nucleophilic aromatic substitution: Experimental and theoretical exploration of a multistep mechanism", *Chem.-Eur. J.* 2008, 14, 3954-3960 e) Nielson, K. D.; van Duin, A.C.T.; Osgaard, J.; Deng, W. Q.; Goddard, W.A. "Development of the ReaxFF reactive force field for describing transition metal catalyzed reactions, with application to the initial stages of the catalytic formation of carbon nanotubes", *J. Phys. Chem. A* 2005, 109, 493-499.
- 37 Dennison, D.M. "On the analysis of certain molecular spectra", *Philos. Mag.* 1926, 1, 195-218.

- 38 van Duin, A.C.T.; Zeiri, Y.; Dubnikova, F.; Kosloff, R.; Goddard, W.A. III, "Atomistic-Scale Simulations of the Initial Chemical Events in the Thermal Initiation of Triacetoneperoxide", *J. Am. Chem. Soc.* 2005, 127, 11053-11062.
- 39 Zhu, R.; Janetzko, F.; Zhang, Y.; van Duin, A.C.T.; Goddard, W.A. III; Salahub, D.R. "Characterization of the active site of yeast RNA polymerase II by DFT and ReaxFF calculations", *Theor. Chem. Acc.* 2008, 120, 479-489.
- 40 Madarász A.; Berta, D.; Paton, R.S. "Development of a true transition state force field from quantum mechanical calculations", *J. Chem. Theory Comput.* 2016, 12, 1833-1844.
- 41 Norrby, P.-O. "Selectivity in asymmetric synthesis from QM-guided molecular mechanics", *J. Mol. Struct. Theochem* 2000, 506, 9-16.
- 42 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.
- 43 Hehre, W.J. "Ab initio molecular orbital theory", *Acc. Chem. Res.* 1976, 9, 399-406.
- 44 Wagner, J.P.; Schreiner, P.R. "London Dispersion in Molecular Chemistry—Reconsidering Steric Effects", *Angew. Chem. Int. Ed.* 2015, 54, 12274-12296.
- 45 Grimme, S. "Density functional theory with London dispersion corrections", *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2011, 1, 211-228.
- 46 Kolb, H.C.; Andersson, P.G.; Sharpless, K.B. "Toward an Understanding of the High Enantioselectivity in the Osmium-Catalyzed Asymmetric Dihydroxylation (AD). 1. Kinetics", *J. Am. Chem. Soc.* 1994, 116, 1278-1291.
- 47 Norrby, P.-O.; Kolb, H.C.; Sharpless, K.B. "Toward an Understanding of the High Enantioselectivity in the Osmium Catalyzed Asymmetric Dihydroxylation (AD). 2. A Qualitative Molecular Mechanics Approach" *J. Am. Chem. Soc.* 1994, 116, 8410-8478
- 48 a) 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 b) Peña-Cabrera, E.; Norrby, P.-O.; Sjögren, M.; Vitagliano, A.; De Felice, V.; Oslob, J.; Ishii, S.; 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.
- 49 Oslob, J.D.; Åkermark, B.; Helquist, P.; Norrby, P.-O. "Steric Influences on the Selectivity in Palladium-Catalyzed Allylation" *Organometallics* 1997, 16, 3015-3021.
- 50 a) 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 b) Hwang, M.-J.; Stockfisch, T.P.; Hagler, A.T. "Derivation of Class II Force Fields. 2. Derivation and Characterization of a Class II Force Field, CFF93, for the Alkyl Functional Group and Alkane Molecules", *J. Am. Chem. Soc.* 1994, 116, 2515-2525.
- 51 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.
- 52 Dasgupta, S.; Goddard, W.A. III, "Hessian-biased force fields from combining theory and experiment", *J. Chem. Phys.* 1989, 90, 7207-7215.
- 53 Norrby, P.O.; Liljefors, T. "Automated Molecular Mechanics Parameterization with Simultaneous Utilization of Experimental and Quantum Chemical Data", *J. Comput. Chem.* 1998, 19, 1146-1166.
- 54 Limé, E.; Norrby, P.-O. "Improving the Q2MM Method for Transition State Force Field Modeling", *J. Comput. Chem.* 2015, 36, 244-250.
- 55 Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Gaussian, Inc., Wallingford CT, 2009.
- 56 Schrödinger Release 2017-1: Jaguar, Schrödinger, LLC, New York, NY, 2017.
- 57 Ponder, J.W.; Case, D.A. "Force Fields for Protein Simulations", *Adv. Protein Chem.* 2003, 66, 27-85.
- 58 a) 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 b) Allinger, N.L.; Zhou, X.; Bergsma, J. "Molecular mechanics parameters", *J. Mol. Struct.: THEOCHEM* 1994, 312, 69-83.
- 59 Schrödinger Release 2017-1: MacroModel, Schrödinger, LLC, New York, NY, 2017.
- 60 Fey, N. J. "Organometallic molecular modelling—the computational chemistry of metallocenes: a review" *J. Chem. Technol. Biotechnol.* 1999, 74, 852-862.
- 61 Norrby, P.-O. "Recipe for an organometallic force field", in *Computational Organometallic Chemistry*, Cundari, T. Ed. Marcel Dekker, Inc: New York, 2001, pp 7-37.
- 62 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.
- 63 a) 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 b) 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.
- 64 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.
- 65 Becke, A.D. "Density-functional thermochemistry. III. The role of exact exchange", *J. Chem. Phys.* 1993, 98, 5648-5652.
- 66 Yamakawa, M.; Noyori, R. "An *ab initio* molecular orbital study on the amino alcohol-promoted reaction of dialkylzinc and aldehydes", *J. Am. Chem. Soc.* 1995, 117, 6327-6335.
- 67 Rasmussen, T.; Norrby, P.-O. "Characterization of New Six-Membered Transition States of the Amino-Alcohol Promoted Addition of Dialkyl Zinc to Aldehydes", *J. Am. Chem. Soc.* 2001, 123, 2464-2465.
- 68 Rasmussen, T.; Norrby, P.-O. "Modeling the Stereoselectivity of the β -Amino Alcohol Promoted Addition of Dialkylzinc to Aldehydes", *J. Am. Chem. Soc.* 2003, 125, 5130-5138.

- 69 Brandt, P.; Norrby, P.-O.; Martin, I.; Rein, T. "A Quantum Chemical Exploration of the Horner-Wadsworth-Emmons Reaction", *J. Org. Chem.* 1998, 63, 1280-1289.
- 70 a) 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 b) 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.
- 71 Leach, A.R. *Molecular Modelling: principles and applications*. Prentice Hall: Harlow, England, 2001
- 72 *Computational Biochemistry and Biophysics*. Becker, O.M.; Mackerell, A.D. Jr.; Roux, B.; Watanabe, M. Eds. Marcel Dekker, Inc: New York, 2001.
- 73 Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. 2nd ed. Springer: New York, 2002.
- 74 Lopes, P.E.M.; Guvench, O.; Mackerell, A.D. "Current Status of Protein Force Fields for Molecular Dynamics", *Methods Mol. Biol.* 2015, 1215, 47-71.
- 75 Warshel, A.; Levitt, M. "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme", *J. Mol. Bio.* 1976, 103, 227-249.
- 76 van der Kamp, M.W.; Mulholland, A.J. "Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology", *Biochemistry* 2013, 52, 2708-2728.
- 77 Wang, C.; Li, W.; Ren, J.; Fang, J.; Ke, H.; Gong, W.; Feng, W.; Wang, C.-C. "Structural Insights into the Redox-Regulated Dynamic Conformations of Human Protein Disulfide Isomerase", *Antioxidants & Redox Signaling* 2013, 19, 36-45.
- 78 Rydberg, P.; Olsen, L.; Norrby, P.-O.; Ryde, U. "General Transition-State Force Field for Cytochrome P450 Hydroxylation", *J. Chem. Theory Comput.* 2007, 3, 1765-1773
- 79 Rydberg, P.; Sigfridsson, E.; Ryde, U. "On the role of the axial ligand in heme proteins: a theoretical study", *J. Biol. Inorg. Chem.* 2004, 9, 203-223.
- 80 Bertz, R.J.; Granneman, G.R. "Use of *In Vitro* and *In Vivo* Data to Estimate the Likelihood of Metabolic Pharmacokinetic Interactions", *Clin. Pharmacokinet.* 1997, 32, 210-258.
- 81 Evans, W.E.; Relling, M.V. "Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics", *Science* 1999, 286, 487-491.
- 82 D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman, AMBER 2016, University of California, San Francisco, 2016.
- 83 Rydberg, P.; Hansen, S.M.; Kongsted, J.; Norrby, P.-O.; Olsen, L.; Ryde, U. "Transition-State Docking of Flunitrazepam and Progesterone in Cytochrome P450", *J. Chem. Theory Comput.* 2008, 4, 673-681.
- 84 Jones, G.; Willet, P.; Glen, R.C.; Leach, A.R.; Taylor, R. "Development and validation of a genetic algorithm for flexible docking", *J. Mol. Bio.* 1997, 267, 727-748.
- 85 Cruciani, G.; Carosati, E.; De Boeck, R.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. "MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist" *J. Med. Chem.* 2005, 48, 6970-6979.
- 86 Oslob, J.D.; Åkermark, B.; Helquist, P.; Norrby, P.-O. "Steric Influences on the Selectivity in Palladium-Catalyzed Allylation", *Organometallics* 1997, 16, 3015-3021.
- 87 For example: a) Houk, K.N.; Liu, F. "Holy Grails for Computational Organic Chemistry and Biochemistry", *Acc. Chem. Res.* 2017, 50, 539-543 b) Hammes-Schiffer, S. "Catalysts by Design: The Power of Theory", *Acc. Chem. Res.* 2017, 50, 561-566 c) Poree, C.; Schoenebeck, F. "A Holy Grail in Chemistry: Computational Catalyst Design: Feasible or Fiction?", *Acc. Chem. Res.* 2017, 50, 605-608, and other contributions in the special issue "Holy Grails in Chemistry".