Transition State Force Field for the Asymmetric Redox Relay Heck Reaction

Anthony R. Rosales[†], Sean P. Ross[‖], Paul Helquist[†], Per-Ola Norrby[‡], Matthew S. Sigman[‖], Olaf Wiest[*,†,§]

[†]Department of Chemistry and Biochemistry, University of Notre Dame, IN 46556, USA

[‡]Data Science and Modelling, Pharmaceutical Sciences, R&D, AstraZeneca Gothenburg, SE-43183 Mölndal, Sweden.

[‖]Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States

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

Abstract

A transition state force field (TSFF) was developed using the quantum-guided molecular mechanics (Q2MM) method to describe the stereodetermining migratory insertion step of the enantioselective redox-relay Heck reaction for a range of multisubstituted alkenes. We show that the TSFF is highly predictive through an external validation of the TSFF against 151 experimentally determined stereoselectivities resulting in an R^2 of 0.89 and MUE of 1.8 kJ/mol. In addition, limitations in the underlying force field were identified by comparison of the TSFF results to DFT level calculations. A novel application of the TSFF was demonstrated for 31 cases where the enantiomer predicted by the TSFF differed from the originally published values. Experimental determination of the absolute configuration demonstrated that the computational predictions were accurate, suggesting that TSFFs can be used for the rapid prediction of the absolute stereochemistry for a class of reactions. Finally, a virtual ligand screen was conducted utilizing both the TSFF and a simple molecular correlation method. Both methods were similarly predictive, but the TSFF was able to show greater utility through transferability, speed, and interpretability.
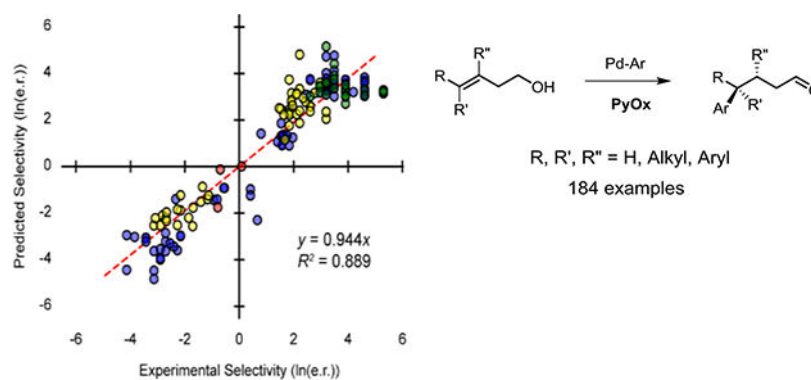
Graphical Abstract

Introduction

A long-standing goal of computational chemistry is to provide fast and easy-to-use tools to accelerate the development of new reactions. This is especially true in asymmetric catalysis, where an underlying understanding of the enantiodetermining step is critical to ligand selection, design, and optimization. The rapid prediction of enantioselectivity of a given substrate-ligand combination has been promised by the computational community for some time, but only recently have advancements been reported to allow this goal to be reached. A common approach in this area is the use of DFT methods that rationalize and ultimately predict stereoselectivity of catalytic reactions.[1–3] However, due to their computational cost and often the complexity of modern reactions, this type of DFT approach is not applicable to evaluating large data sets. Additionally, these methods are often limited to only relatively small subsets of conformations,[4] which can lead to deviations from experimental results if relevant transition state conformations are not found in the initial search.

An emerging alternative approach is to use correlation methods where experimental results for yields or selectivities of often designed datasets including ligand/substrate combinations are correlated with molecular features derived from surrogate structures of both reactants and catalysts.[5] The proper selection of parameters and curation of experimental data, for example in the form of designed or "clean" datasets, are essential for the development of statistical models that can predict out-of-sample cases.[6] Related approaches include more recent machine-learning (ML) methods,[7] which require rather extensive experimental data sets.[8] As with many ML-derived correlation methods, the interpretability of the models is often a challenge.[5, 9]

An alternative strategy for the rapid prediction of stereoselectivities is to model the transition state as a minimum using transition state force fields (TSFFs) derived by the quantum guided molecular mechanics (Q2MM) method.[10–11] Although Q2MM is also a fitting procedure, there are several key differences to the correlation strategies discussed above. First, no experimental data are used in the fitting procedure, making the method truly predictive. Second, predictions are made using information from the physically relevant, stereo-determining transition states rather than surrogate structures. We have previously used TSFFs to rapidly predict the ratios of stereoisomers for a wide range of reactions with mean unsigned errors (MUE) of 2–3 kJ/mol.[12] In Q2MM, suitable electronic structure calculations

of the stereo-determining transition structures (TS) of interest are used to determine information such as geometries, Hessian matrix elements, or partial charges that are used to fit a classical force field. This force field is then used to calculate the conformational ensembles for a set of diastereomeric transition states by Monte-Carlo sampling.[13–14] The Boltzmann averaged free energies of this ensemble provides not only the relative Gibbs free energies ($G^{\ddagger}$), but can also give structural insights into the origin of the stereoselection. Recent work in this field has been focused on developing easy-to-use software for the generation of TSFFs (Q2MM) and evaluating reaction conditions (CatVS) that is accessible to experimental researchers.[10–12]

In this context and related to our ongoing interest in enantioselective redox-relay Heck reactions, we deployed our TSFF approach to understand and predict enantioselective outcomes for this broad reaction class. The reported examples of this reaction use, in contrast to the classical Heck reaction,[15] generally electronically unbiased alkenes (Scheme 1).[16–21] We have previously studied the mechanism both experimentally and through traditional DFT transition state analysis.[22–24] These studies have revealed that: 1) migratory insertion of the aryl–Pd species generated from the arylboronic acid[25–26] into the coordinated alkene is both the rate- and stereo-determining step, 2) the substrate remains bound to the Pd-catalyst through sequential migratory insertion/β-hydride elimination (chain-walking) until base-assisted reduction of the metal and product formation, and as a consequence 3) the reaction provides stereocontrol at both carbons of the alkene.

Driven by the wide and evolving scope of the reaction in terms of the alkene, nucleophile, and redox acceptor (the functional group attached to the alkene that promotes termination of chain-walking) and the utility of the products, herein we describe the use of the Q2MM method to develop a TSFF to predict the stereoselectivity of the redox-relay Heck reaction. Specifically, we describe the development and application of a TSFF describing the migratory insertion TS. The TSFF was validated using published results on five classes of substrates, resulting in a larger data set than is typically explored with Q2MM. This analysis revealed that for one distinct class of substrates, indole coupling partners, the predicted enantiomer was inconsistent with the assignment of configuration in the initial report. This prompted us to explore this unexpected reversal of enantioselectivity through a combination of DFT methods and experimental studies. This ultimately exposed that the experimentally reported enantiomer was incorrect, which highlights the ability of the TSFF to predict the correct enantiomer of the product. Finally, the TSFF is applied to virtual screening of ligands providing, for the first time, a direct comparison between predictions made using correlative methods and those from TSFFs.

Methods

The parameterization of the TSFF followed the procedures outlined in the literature[10, 12] using Q2MM.[3] The reference data used to train the TSFF were obtained by optimizing simplified models of the transition structures (Figure S1 in Supplementary Information(SI)) in G09[27] using M06-GD3 with a Lanl2DZ basis set for palladium and 6–31+G* for all other atoms. MM3*, as implemented in MacroModel,[28] was used as the underlying functional form.[29] All reference structures and selected transition structures for the full system were

optimized and characterized as having a single imaginary frequency corresponding to the reaction coordinate. Partial charges were obtained using the CHELPG method, with an atomic radius of 2.4 Å for palladium. External validation to experimental selectivities used CatVS[10] to automatically set conformational search parameters. The conformational search and subsequent minimizations were also carried out in MacroModel, with the predicted selectivity calculated from the Boltzmann averaged conformations. Selectivities are referenced to the (*R*)-stereoisomer of the newly formed chiral center, in that if the major stereoisomer has the (*S*)-configuration, the selectivity is reported as negative (alkene stereoisomers generally lead to opposite product configurations). M06-GD3/6–311++G** was used to calculate distortion-interaction energies.[23] The interaction energies were calculated between the alkene substituent and oxazoline moieties using the TS-optimized geometries.

Results and Discussion

TSFF Development.

The TSFF was fitted for the migratory insertion of an aryl nucleophile across the alkene. A training set of 24 transition structure models (Figure S1 in SI) was obtained by optimizations of 11 different combinations of PyrOx ligands and substrates using the level of theory described above. These TSs are simplified model systems where each combination describes a particular interaction or distortion independent of remote substituents. Structural, electrostatic, and force constant information from the training set was used to fit the TSFF parameters following the Q2MM workflow.[3, 10, 12] The TSFF was validated against the training set by comparing normal modes and geometries obtained from DFT level calculations to the results of the TSFF (Figures S3–S6 in SI). These comparisons showed excellent correlation between the TSFF and DFT calculated data, with slope and R^2 values near unity (Figure S3–6 in SI). The largest deviations between the TSFF and the DFT training set are due to the underlying MM3*. Therefore, the errors due to the TSFF parameters are within the range of all other errors, and the TSFF was used to predict experimental selectivities.

Validation to experimental selectivities.

Once built, the goal of TSFFs is to provide a reliable and rapid tool for the prediction of enantiomeric ratios from a given set of substrates and ligands. To validate this tool, 183 experimental selectivities from the literature[16–21] were determined using the TSFF. The experimental results include various substituted unbiased alkenes with remote functional groups (e.g., alcohols or π-EWG) and are summarized in Figure 1. This experimental validation set is significantly larger than the ones used for previously published TSFFs, which typically contained 20–40 experimentally determined selectivities.[13–14, 30]

The performance of the TSFF for the arylation of 76 internal di- and trisubstituted alkenols resulted in an MUE of 1.8 kJ/mol (blue Figure 1).[17–18] In the case of disubstituted alkenes, the TSFF predicts the major enantiomer of the major regioisomer (arylation of the alkene carbon distal to the functional group). Importantly, while this specific type of nucleophile-substrate combination is included in the training set, these are still true predictions because

the training set is exclusively computationally derived. Lastly, these results are comparable to previous TSFFs[10, 12] and are sufficiently accurate to be useful for high throughput predictions of stereoselectivity.[11]

A limitation of the TSFF is that quinoline-derived oxazoline (QuinOx) ligands predicted the opposite enantiomer (Scheme 2). To determine the cause for this discrepancy, the result with the largest error was analyzed by DFT. The lowest energy DFT transition structures show a high degree of distortion of the TS that is not captured by the TSFF. Most notably, the Pd–N_{Pyr} and forming $C_{Aryl}···C_{Alkene}$ distances (blue in Figure 1) are 0.11–0.16 and 0.09–0.13 Å longer, respectively (Table S3 in Supplementary Information). The enantioselectivity predicted by DFT is 2.5 kJ/mol (48% ee), in agreement with the experimental results (32% ee) but in contrast to the one predicted by the TSFF of −5.8 kJ/mol (83% ee). This shows that the TSFF is limited to ligands that do not introduce a significant degree of distortion relative to the training set.

We next explored whether the TSFF would accurately predict the selectivity of substrate classes not present in the training set. Specifically, we predicted the selectivity of 26 substrates containing remote π-electron withdrawing groups (green in Figure 1). This resulted in an MUE of 2.1 kJ/mol between predicted and experimental selectivities, which is similar to the arylation of di- and trisubstituted alkenes (1.8 kJ/mol). In contrast to the di- and trisubstituted alkenes, this data set contains few examples of reactions with low enantioselectivity. While this uniformly high degree of enantioselectivity is ideal for synthetic purposes, it limits our ability to access the predictive capabilities of this class of reactions. These results demonstrate that the TSFF is capable of predicting the enantioselectivity of di- and trisubstituted alkenes regardless of the nature of the redox acceptor. This is consistent with the proposed mechanism in which the selectivity is governed by the local TS environment (i.e., nonbonded/bonded interactions about the square plane).[22]

Two sets of substrates show poor correlation between the TSFF predictions and the experimental results (yellow and red in Figure 1). Although these subsets are significantly smaller than the other validation sets, and in many cases have only modest selectivities, such deviations have historically been important to understand the shortcomings of force fields. Therefore, we sought to analyze these subsets in detail with the goal of aiding the development of more accurate methods.[31–32]

We began with the arylation of 1,1-disubstituted alkenols (Scheme 3). Our previous studies have shown that the enantioselectivity of the migratory insertion is dictated primarily by the steric difference between the cis-oriented alkyl and hydrogen substituents on the alkene.[16] However, this analysis does not translate to 1,1-disubstitued alkenes. Therefore, to determine the source of the selectivity with this class of alkenes, the reaction with the largest deviation between the predicted and experimentally determined enantioselectivity was studied by DFT. The DFT calculations (M06-GD3/Lanl2DZ(Pd)/6–31+G*//SDD(Pd)/6–311++G** + SMD(DMF)) revealed a ΔG‡ of 7.3 kJ/mol (91% ee), which agrees with the experimentally observed 9.4 kJ/mol (96% ee) and in contrast to the −3.3 kJ/mol (−59% ee) predicted by

TSFF. Although the energies are not correct in this case, the TSFF can still be a useful conformational search tool in analogy to the previously studied Mukaiyama aldol reaction.[33]

Unlike alkenes monosubstituted at the 1-position or disubstituted and the 1- and 2-positions, the enantioselectivity for the 1,1-disubstituted alkenols is governed by the same *cis-/trans*-isomer of the square planar palladium complex where the major paths to both enantiomers of the product have the transferring aryl group coordinated *trans* to the oxazoline moiety. Therefore, to delineate the source of the selectivity and the origin of the disagreement between the DFT and TSFFs results, a distortion-interaction analysis[23] at the M06-GD3/6–311++G** level of theory was performed (Figure 2).

The difference in distortion energies of the alkene and ligand components for both **TS^R** and **TS^S** are small, leading to a net 1.3 kJ/mol difference favoring the (*R*)-enantiomer. The interaction energies of the complex show that the presence of stabilizing non-covalent interactions between the aromatic substituent of the alkene and the C–H of the phenyl substituent on the ligand. In the case of **TS^R**, there are two such interactions (Figure 2, bottom left), while for **TS^S** there is only one interaction (Figure 2, bottom right). This results in the (*R*)-enantiomer being favored by 4.8 kJ/mol. These factors together favor the (*R*)-enantiomer by 6.1 kJ/mol, in agreement with DFT and experimental results.[22, 24, 34]

An analogous interaction analysis was performed using the TSFF. However, because the ground state alkenol and ligand structures and the TS are described using different parameter sets, the distortion analysis cannot be performed in the exact same manner. The interaction energies using the TSFF systematically under predicts the aromatic interaction energies by ~10 kJ/mol, while the alkyl-alkyl interactions are under predicted by 2 kJ/mol (Scheme 2 bottom, parenthesis). This results in the (*S*)-enantiomer being preferred by 5.2 kJ/mol. The under-prediction of the degree of interaction energy imparted by the interactions with the aryl substituents results in a 9.4 kJ/mol discrepancy between the TSFF and DFT result. This demonstrates that the systematic error (MUE = 7.8 kJ/mol) of the 1,1-disubstituted alkenols is due to an inaccuracy of the MM3* FF in describing the aromatic interactions.

These results are further supported by two benchmarking studies[35–36] that showed large deviations in non-bonded aromatic interactions when using MM3*. Paton and Goodman found MUEs of 2.7 and 1.3 kJ/mol for the benzene/methane and stacked benzene dimers, respectively.[35] Additionally, complexes with primarily dispersive interactions (6 of the 8 contained aromatic compound in the study) were associated with an RMSD of 0.61 Å. Sherrill, *et al.* found these errors are greater when the structures are not minima on the potential energy surface.[36] They concluded atom-centered electrostatic models in traditional FFs are unable to describe π-interactions, especially those that are not at low-energy conformations. This demonstrates the systematic error is not due to the TSFF parameters, but rather the underlying functional form of MM3*. It should be noted that stereoselectivity predictions for reactions that have either a dialkyl substituted alkenol substrate (red circles) or a ligand with an sp^3 hybridized carbon bound to C4 of the oxazoline are correct. This further supports the hypothesis that this error is associated with a substrate-ligand combination due to a specific nonbonded interaction.

The second reaction that showed significant discrepancies (MUE 8.2 kJ/mol) between the TSFF and experimental results[19] is the C3-alkylation of indoles (yellow circles in Figure 1). Notably, there is a set of outliers in the upper left quadrant of Figure 1, which all arise from the use of an *R*-CH$_2$Naph substituent on the oxazoline. The placement of these outliers suggests that the degree of selectivity is accurately predicted, however with the opposite absolute configuration. Furthermore, this subset was reported to have the same absolute configuration as the structurally similar (*S*)-CH$_2$Ph PyrOx ligand. This not only raises doubt about the reported absolute stereochemistry, but more importantly has the intriguing implication that, if shown to be correct, TSFFs could be used to rapidly "proofread" absolute stereochemical assignments.

To assess the accuracy of the TSFF result, we computed the TSs of the *R*-CH$_2$Naph (**L1**), (*S*)-CH$_2$Ph (**L2**), and (*R*)-Ph (**L3**) PyrOx ligands by DFT at the M06-GD3 level of theory (Table 1). For **L1**, the DFT and TSFF results both predict the (*R*)-enantiomer, while the literature reports the (*S*)-enantiomer. Contrastingly, for both **L2** and **L3** the DFT and TSFF agree with the literature report. These results suggest that in the case of **L1** the enantiomer was assigned incorrectly in the initial report. In the initial report, the absolute configuration was not experimentally determined. Instead, it was demonstrated that the reaction proceeds through a *syn*-nucleopalladation pathway by comparing the enantioselectivity between arylation reactions with a 3-H and 3-BPin indole substrates (Scheme 4a). From this result, it was inferred that the reaction proceeds with the same sense of enantioselectivity as the arylation with arylboronic acids. However, this analysis does not unambiguously assign the absolute configuration. Therefore, we set out to confirm the absolute configuration of the C3-alkylated products using **L1** in order to determine the absolute configuration.

In order to determine the absolute configuration of these products, we targeted compounds we believed would result in X-ray quality single crystals (Scheme 4). Towards this end, we performed the dehydrogenative Heck arylation[20] of trisubstituted allylic alcohol **1** with indole **2** using the (*S*)-enantiomer of **L1** to give the aldehyde product, which was then converted into the corresponding tetrachlorophthalimide **3**. Initial attempts to determine the absolute stereochemistry of **3** using X-ray crystallography were unfortunately not successful because **3** consistently crystallized as a racemate regardless of the crystallization conditions used. We therefore determined the absolute stereochemistry by comparison of the experimental and predicted ECD spectrum (Figure 3, for full computational details, see Supplementary Information).

In agreement with earlier benchmarking studies,[37] our results show that the overall features of the predicted ECD spectra are similar between different functionals, but slight shifts between the wavelengths of the maxima are evident. Nevertheless, the results unequivocally confirm the absolute stereochemistry as (*R*), in agreement with the computational results but in contrast to the previous assignment of the absolute configuration.[19]

Overall, when the 1,1 disubstituted alkyl styrenes are excluded (for the reasons discussed above) and using the corrected stereochemistries, the TSFF shows an excellent correlation to the 151 experimental results (Figure 4). Lastly, the total MUE of 1.8 kJ/mol over a wide

range of selectivities is comparable to previously developed TSFFs using Q2MM having MUEs of 2.5–2.7 kJ/mol.[10, 12]

Virtual Screening of Ligands.

One important application of TSFFs, correlation methods, and machine learning is the virtual screening of chemical libraries.[38] As a proof of principle, we applied the TSFF to a virtual ligand screen for reactions that showed low levels of enantioselectivity.

The first class of substrates studied were 1,1-dialkyl-substituted alkenols (red in Figure 4). Unfortunately, when this class of substrates was screened against the ligand library (53 combinations), no significant improvements were found. This confirms the experimental finding that 1,1-disubstituted alkenes are unlikely to undergo highly selective arylations using the current ligand design.

The second reaction class studied is the C3-alkylation of indoles, which consistently performed below 90% ee, whereas the ethyl-substituted analogs resulted in >90% ee (Figure 5, grey).[19] Sigman and coworkers observed a correlation between the NBO of the oxazoline nitrogen and the enantioselectivity based on an initial ligand screen (Figure 5, grey circles). Using this linear free energy relationship (LFER), they virtually evaluated a number of ligands. Based on this virtual screen, a set of eight ligands were synthesized resulting in an increase in the selectivity from 74% ee to 86% ee (Figure 5, grey **x**). Importantly, in light of the reassignment discussed, this LFER predicts the magnitude of the selectivity, with the optimal catalysts giving the enantiomer of the product compared to the ligands screened. This highlights that utilization of achiral parameters such as NBOs can be useful for prediction of the magnitude of selectivity, but not necessarily the absolute stereochemical outcome.

We saw this as an opportunity to directly compare, for the first time, the strengths and weaknesses of TSFF and correlation-based predictions. Therefore, we virtually screened the same eight ligands with the TSFF in order to provide a quantitative comparison to a LFER-type correlation method (Figure 5, yellow). The magnitudes of the selectivities predicted from the LFER (Figure 5, grey) and TSFF (Figure 5, yellow) show that both methods provide accurate predictions within an absolute error of 2 kJ/mol for this data set, except for two datapoints in both the LFER and TSFF data sets that are slightly above this arbitrary limit. The extrapolation data in the LFER set (grey circles) resulted in an MUE of 1.2 kJ/mol. For these same reactions, the TSFF produced a 1.3 kJ/mol MUE. Overall, a 1.2 kJ/mol MUE was achieved for the TSFF data set. This is interesting because the ligands screened contained changes in steric and electronic substitutions, the latter of which is often not described well in FFs with static charge models.

This case study highlights many of the strengths and weaknesses of each method. Although the continued development of the Q2MM code greatly accelerated the fitting of TSFFs, the elucidation of the stereo-determining step of a reaction, followed by calculation of model systems, fitting and validation of the TSFF currently requires more effort than the fitting of a small experimental dataset using correlation approaches, which do not require any mechanistic insight. In turn, predictions from correlation methods, even though they can aid

in the formulation of hypotheses, do not *a priori* provide insights into the physical origin of the predicted selectivity and require chemical intuition (and further experimentation and/or computation such as DFT) as to why a descriptor is correlated with the selectivity. In contrast, a well validated TSFF describes the structural and electronic nature of the TSs and is therefore expected to be more robust towards extrapolation beyond the original training set. Furthermore, this work demonstrates the power of TSFF of predicting the absolute stereochemistry as this information is encoded into the approach. Finally, it is important to emphasize that TSFFs are truly predictive in that they do not use any empirical information in the fitting of the force field. This results in a purely computationally determined model, which fulfills a promise of computational chemistry. Lastly, we consider correlative and TSFF methods to be complementary, with each suited to specific applications. For example, in instances such as the present study, where the mechanistic details are well-understood, the development of TSFFs provides a powerful method that not only predicts the relative selectivity, but is also able to predict the absolute selectivity. Contrastingly, correlative methods do not require detailed knowledge of the mechanistic features, and are typically easier to implement especially during an optimization campaign.

Conclusions

The TSFF for the Pd-catalyzed arylation of a range of electronically unbiased olefins accurately predicts the stereoselectivity with an MUE of 1.8 kJ/mol and an R^2 of 0.889 over the largest validation set yet containing 151 examples covering a wide range of substrates and ligands. It is the first example of a quantitatively accurate Q2MM-derived TSFF for a carbon-carbon bond formation. The physical origin of the deviation of a small subset of outliers, the arylation of styrenes, was traced back to the unsatisfactory treatment of aryl-aryl interactions in the underlying MM3* force field. The accuracy of the TSFF was demonstrated by its ability to predict the correct absolute configuration of a previously missassigned result from the literature. This suggests a new application of TSFFs in the rapid "proofreading" of stereochemical assignments. Lastly, we applied both the structure-based TSFF and a statistical model developed for the same reaction to the virtual screening of ligands, allowing for the first direct comparison between the two complementary methods. The advantages and disadvantages of each approach are discussed in detail.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

REFERENCES

1. Lam Y-H; Grayson MN; Holland MC; Simon A; Houk K, Theory and modeling of asymmetric catalytic reactions. Acc. Chem. Res 2016, 49, 750–762. [PubMed: 26967569]

2. Xu W; Arieno M; Lö H; Huang K; Xie X; Cruchter T; Ma Q; Xi J; Huang B; Wiest O, Metal-templated design: enantioselective hydrogen-bond-driven catalysis requiring only parts-per-million catalyst loading. J. Am. Chem. Soc 2016, 138, 8774–8780. [PubMed: 27336458]

3. github.com/q2mm.

4. Guan Y; Ingman VM; Rooks BJ; Wheeler SE, AARON: An automated reaction optimizer for new catalysts. J. Chem. Theor. Comp 2018, 14, 5249–5261.

5. Reid JP; Sigman MS, Comparing quantitative prediction methods for the discovery of small-molecule chiral catalysts. Nature Rev. Chem 2018, 2, 290–305.

6. Reid JP; Sigman MS, Holistic prediction of enantioselectivity in asymmetric catalysis. Nature 2019, 571, 343–348. [PubMed: 31316193]

7. Zahrt AF; Henle JJ; Rose BT; Wang Y; Darrow WT; Denmark SE, Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. Science 2019, 363, eaau5631. [PubMed: 30655414]

8. Ahneman DT; Estrada JG; Lin S; Dreher SD; Doyle AG, Predicting reaction performance in C–N cross-coupling using machine learning. Science 2018, 360, 186–190. [PubMed: 29449509]

9. Chuang KV; Keiser MJ, Comment on "Predicting reaction performance in C–N cross-coupling using machine learning". Science 2018, 362 (6416), eaat8603. [PubMed: 30442776]

10. Rosales AR; Quinn TR; Wahlers J; Tomberg A; Zhang X; Helquist P; Wiest O; Norrby P-O, Application of Q2MM to predictions in stereoselective synthesis. Chem. Comm 2018, 54, 8294–8311. [PubMed: 29971313]

11. Rosales AR; Wahlers J; Limé E; Meadows RE; Leslie KW; Savin R; Bell F; Hansen E; Helquist P; Munday RH; Wiest O; Norrby PO, Rapid virtual screening of enantioselective catalysts using CatVS. Nature Catalysis 2019, 2, 41–45.

12. Hansen E; Rosales AR; Tutkowski B; Norrby P-O; Wiest O, Prediction of Stereochemistry using Q2MM. Acc. Chem. Res 2016, 49, 996–1005. [PubMed: 27064579]

13. Donoghue PJ; Helquist P; Norrby P-O; Wiest O, Prediction of enantioselectivity in rhodium catalyzed hydrogenations. J. Am. Chem. Soc 2008, 131, 410–411.

14. Donoghue PJ; Helquist P; Norrby P-O; Wiest O, Development of a Q2MM force field for the asymmetric rhodium catalyzed hydrogenation of enamides. J. Chem Theor. Comp 2008, 4, 1313–1323.

15. Heck RF, Palladium-catalyzed reactions of organic halides with olefins. Acc. Chem. Res 1979, 12, 146–151.

16. Werner EW; Mei T-S; Burckle AJ; Sigman MS, Enantioselective Heck arylations of acyclic alkenyl alcohols using a redox-relay strategy. Science 2012, 338, 1455–1458. [PubMed: 23239733]

17. Mei T-S; Patel HH; Sigman MS, Enantioselective construction of remote quaternary stereocentres. Nature 2014, 508, 340. [PubMed: 24717439]

18. Mei T-S; Werner EW; Burckle AJ; Sigman MS, Enantioselective redox-relay oxidative Heck arylations of acyclic alkenyl alcohols using boronic acids. J. Am. Chem. Soc 2013, 135, 6830–6833. [PubMed: 23607624]

19. Zhang C; Santiago CB; Crawford JM; Sigman MS, Enantioselective dehydrogenative Heck arylations of trisubstituted alkenes with indoles to construct quaternary stereocenters. J. Am. Chem. Soc 2015, 137, 15668–15671. [PubMed: 26624236]

20. Zhang C; Santiago CB; Kou L; Sigman MS, Alkenyl carbonyl derivatives in enantioselective redox relay Heck reactions: accessing α, β-unsaturated systems. J. Am. Chem. Soc 2015, 137, 7290–7293. [PubMed: 26030059]

21. Chen Z-M; Hilton MJ; Sigman MS, Palladium-catalyzed enantioselective redox-relay heck arylation of 1, 1-disubstituted homoallylic alcohols. J. Am. Chem. Soc 2016, 138, 11461–11464. [PubMed: 27571167]

22. Xu LH, M. J; Zhang X; Norrby PO; Wu YD; Sigman MS; Wiest O, Mechanism, Reactivity and Selectivity in Palladium-Catalyzed Redox-Relay Heck Arylations of Alkenyl Alcohols. J. Am. Chem. Soc 2014, 136, 1960–1967. [PubMed: 24410393]

23. Hilton MJ; Cheng B; Buckley BR; Xu L; Wiest O; Sigman MS, Relative reactivity of alkenyl alcohols in the palladium-catalyzed redox-relay Heck reaction. Tetrahedron 2015, 71, 6513–6518. [PubMed: 26392640]

24. Hilton MJ; Xu L-P; Norrby P-O; Wu Y-D; Wiest O; Sigman MS, Investigating the nature of palladium chain-walking in the enantioselective redox-relay Heck reaction of alkenyl alcohols. J. Org. Chem 2014, 79, 11841–11850. [PubMed: 25186804]

25. Andappan MM; Nilsson P; von Schenck H; Larhed M, Dioxygen-promoted regioselective oxidative Heck arylations of electron-rich olefins with arylboronic acids. J. Org. Chem 2004, 69, 5212–5218. [PubMed: 15287763]

26. Jung YC; Mishra RK; Yoon CH; Jung KW, Oxygen-promoted Pd (II) catalysis for the coupling of organoboron compounds and olefins. Org. Lett 2003, (13), 2231–2234. [PubMed: 12816416]

27. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M,;, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery J, J. A;, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ Gaussian 09, Revision A1, Wallingford, CT, 2009.

28. Schrödinger Release 2017–3: MacroModel, Schrödinger LLC, New York, NY 2017.

29. Allinger NL; Yuh YH; Lii JH, Molecular mechanics. The MM3 force field for hydrocarbons. 1. J. Am. Chem. Soc 1989, 111, 8551–8566.

30. Limé E; Lundholm MD; Forbes A; Wiest O; Helquist P; Norrby P-O, Stereoselectivity in asymmetric catalysis: the case of ruthenium-catalyzed ketone hydrogenation. J. Chem Theor. Comp 2014, 10, 2427–2435.

31. Hagler AT, Force field development phase II: Relaxation of physics-based criteria… or inclusion of more rigorous physics into the representation of molecular energetics. J. Comp. Aid. Des 2019, 33, 205–264.

32. Dauber-Osguthorpe P; Hagler AT, Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there? J. Comp. Aid. Des 2019, 33, 133–203.

33. Lee JM; Zhang X; Norrby P-O; Helquist P; Wiest O, Stereoselectivity in (Acyloxy) borane-Catalyzed Mukaiyama Aldol Reactions. J. Org. Chem 2016, 81, 5314–5321. [PubMed: 27247023]

34. Hilton MJ Mechanistic Studies of Palladium-Catalyzed Enantioselective Alkene Functionalization Reactions using Classical and Modern Stategies, PhD Thesis, Unversity of Utah, 2017.

35. Paton RS Goodman JM, Hydrogen Bonding and π-Stacking: How Reliable Are Force Fields? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions. J. Chem. Inf. Mod 2009, 49, 944–955.

36. Sherrill CDS, B. G; Sinnokrot MO; Marshall MS; Hohenstein EG; Walker RC; Gould IR, Assessment of Standard Force Field Models Against High-Quality Ab Initio Potential Curves for Prototypes of π-π, CH/π, and SH/π Interactions. J. Comp. Chem 2009, 30, 2187–2193. [PubMed: 19242959]

37. Brklja a Z; Mališ M; Smith DM; Smith A-S, Calculating cd spectra of flexible peptides: An assessment of td-dft functionals. J. Chem Theor. Comp 2014, 10, 3270–3279.

38. Rosales AR; Wahlers J; Limé E; Meadows RE; Leslie KW; Savin R; Bell F; Hansen E; Helquist P; Munday RH; Wiest O; Norrby P-O, Rapid virtual screening of enantioselective catalysts using CatVS. Nature Catalysis 2018, 2, 41–45.
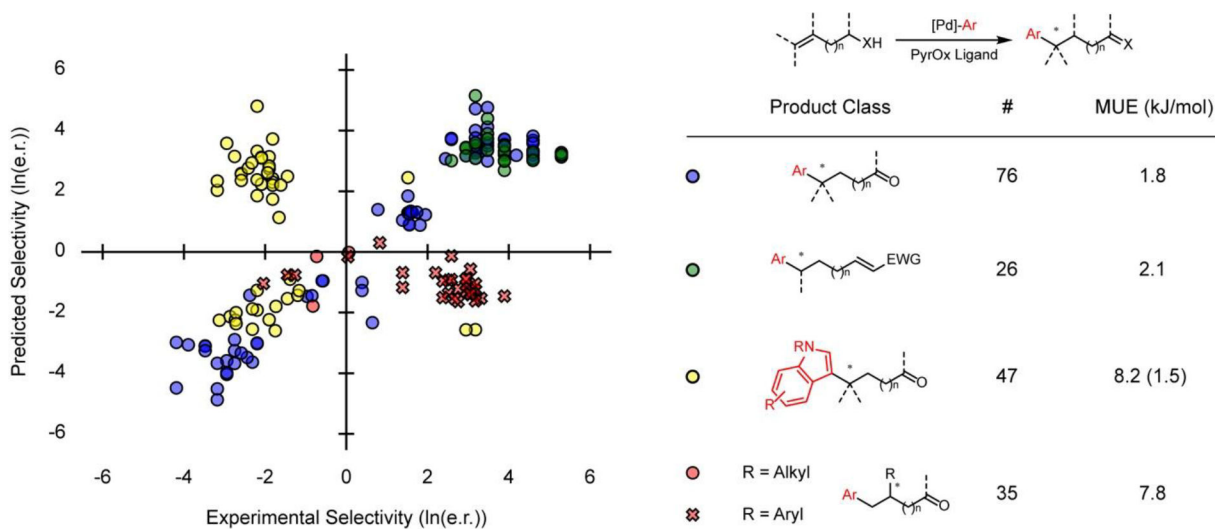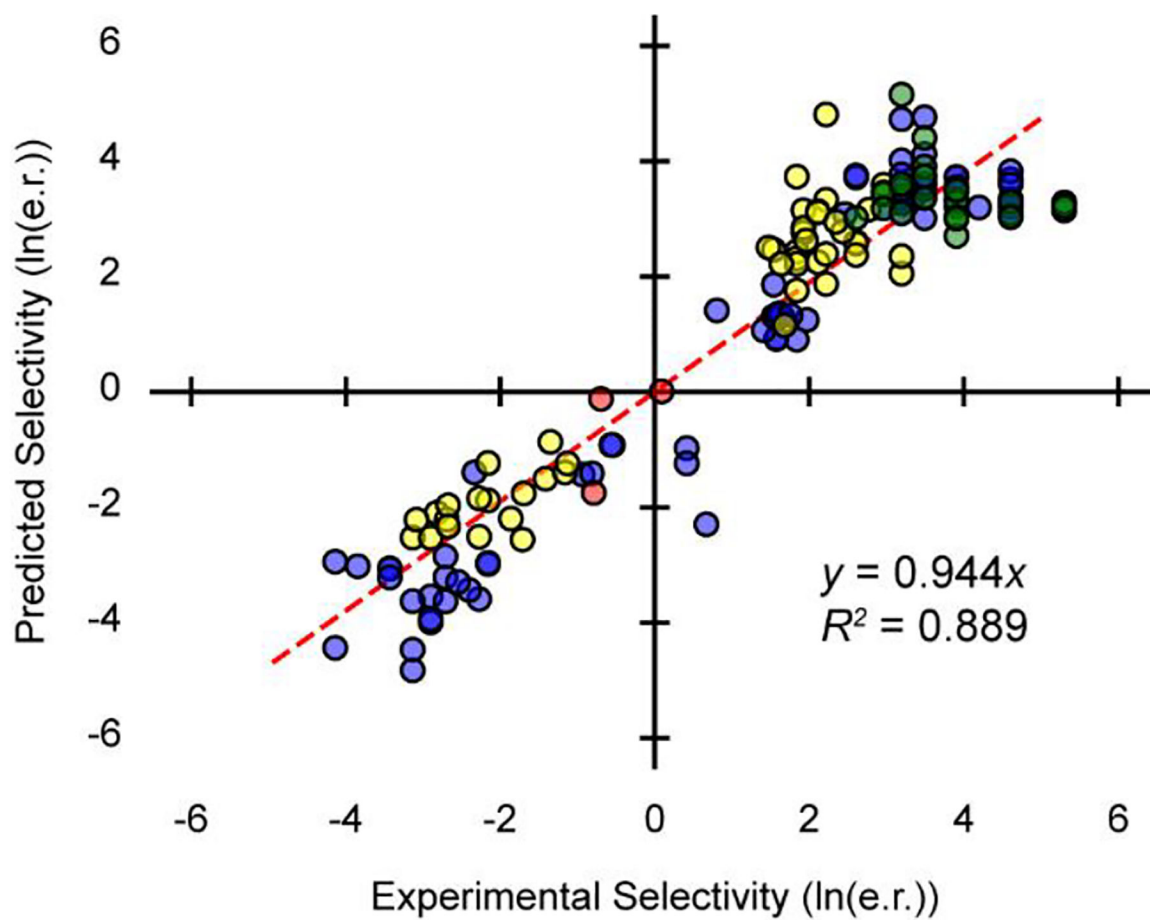
Figure 1:

Comparison of 184 predicted and experimental selectivities. Stereochemistry (*R*/*S*) is indicated by +/− values, respectively. The MUE of the magnitudes of the selectivity, omitting absolute configuration, is included in parentheses.
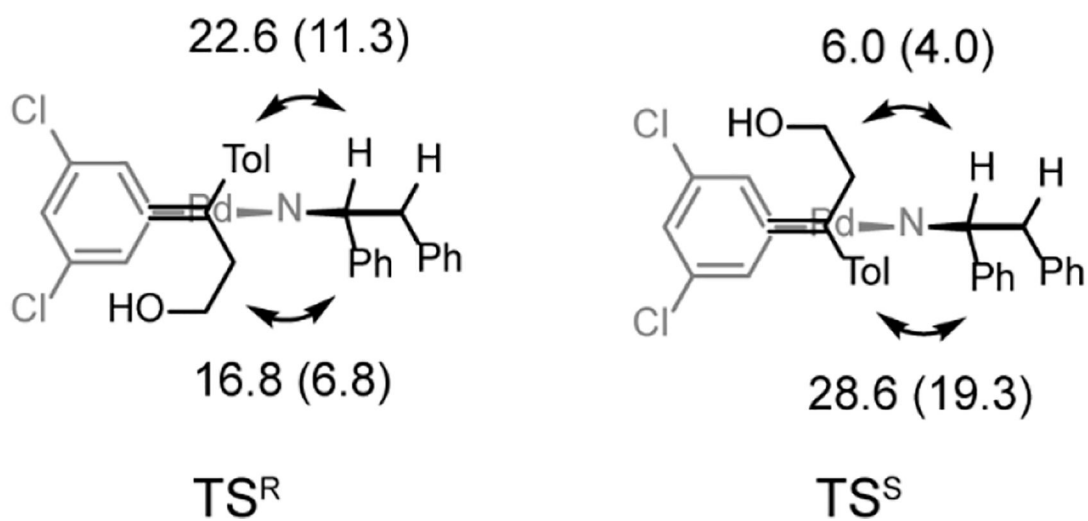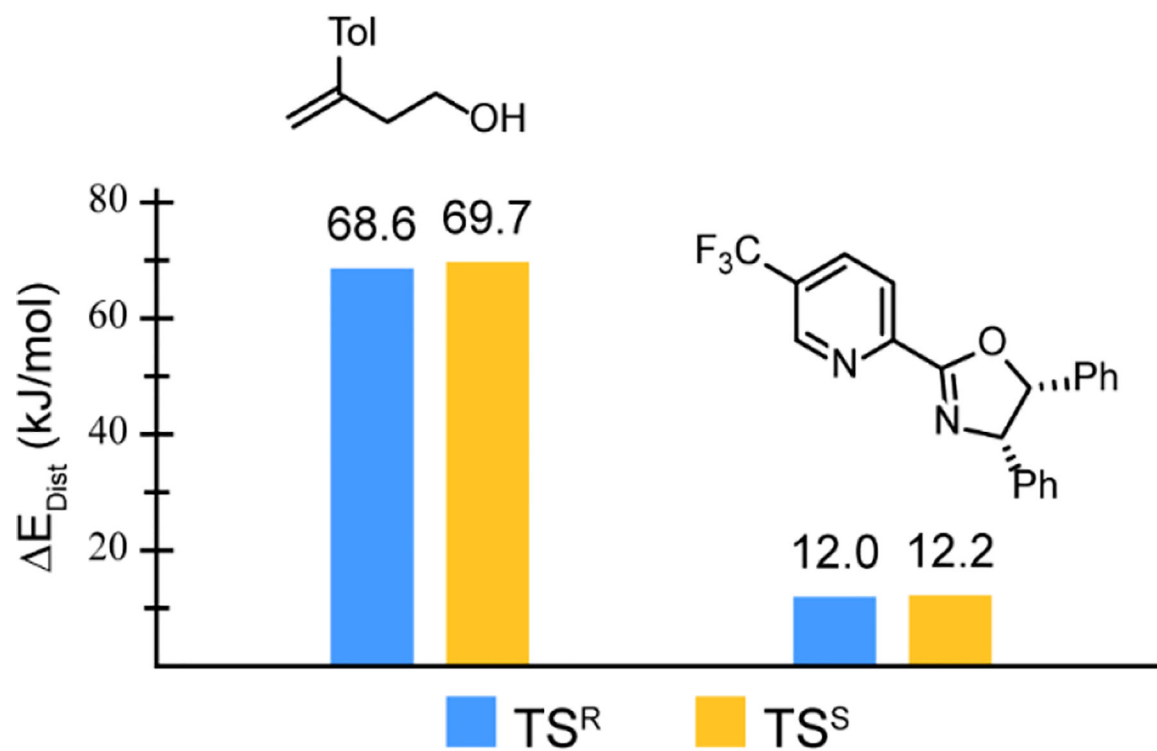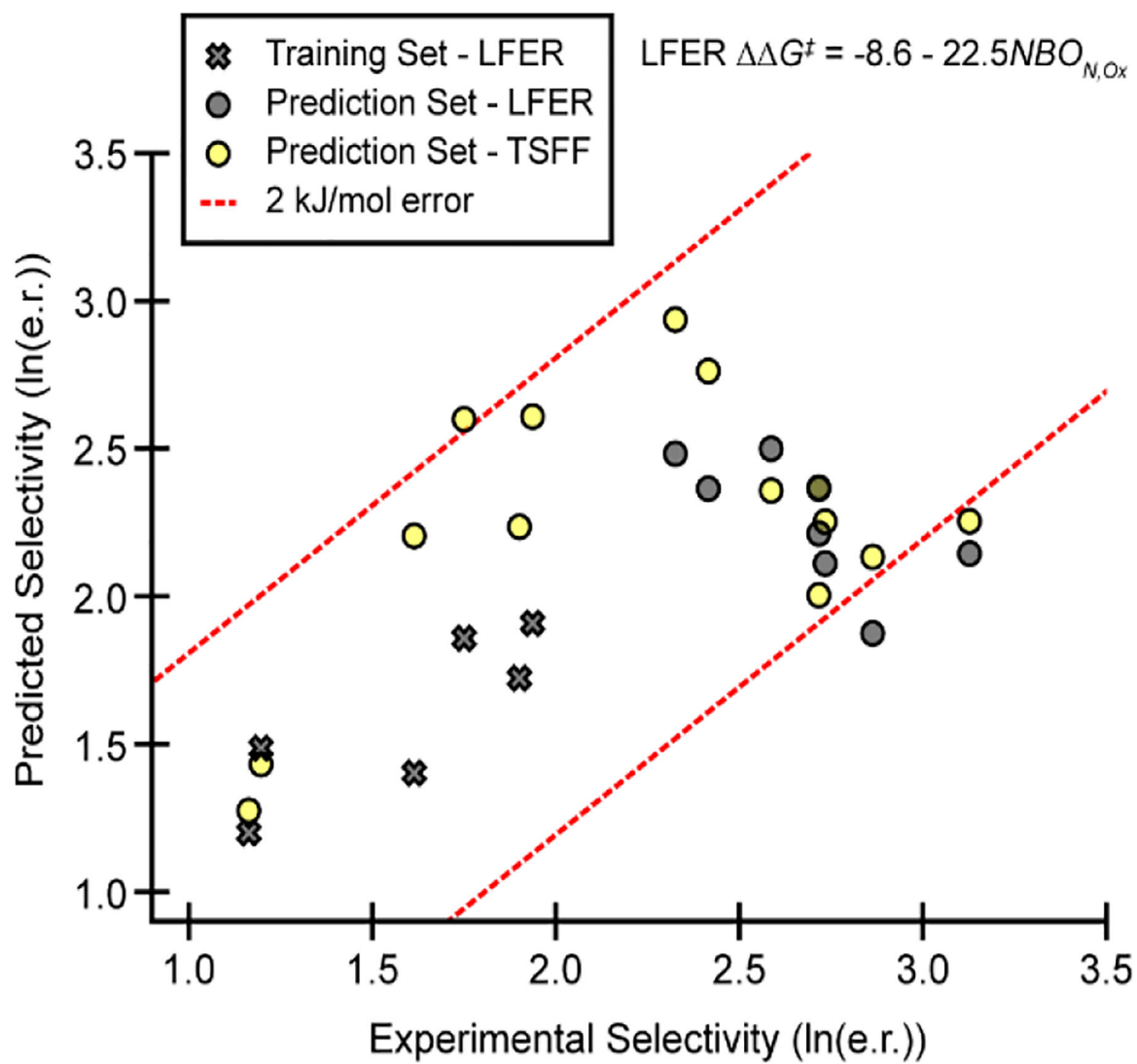
Figure 2:
Distortion (top) and interaction (bottom) energy analysis. Energies (kJ/mol) are reported as M06-GD3/6–311++G** with MM3* values in parentheses
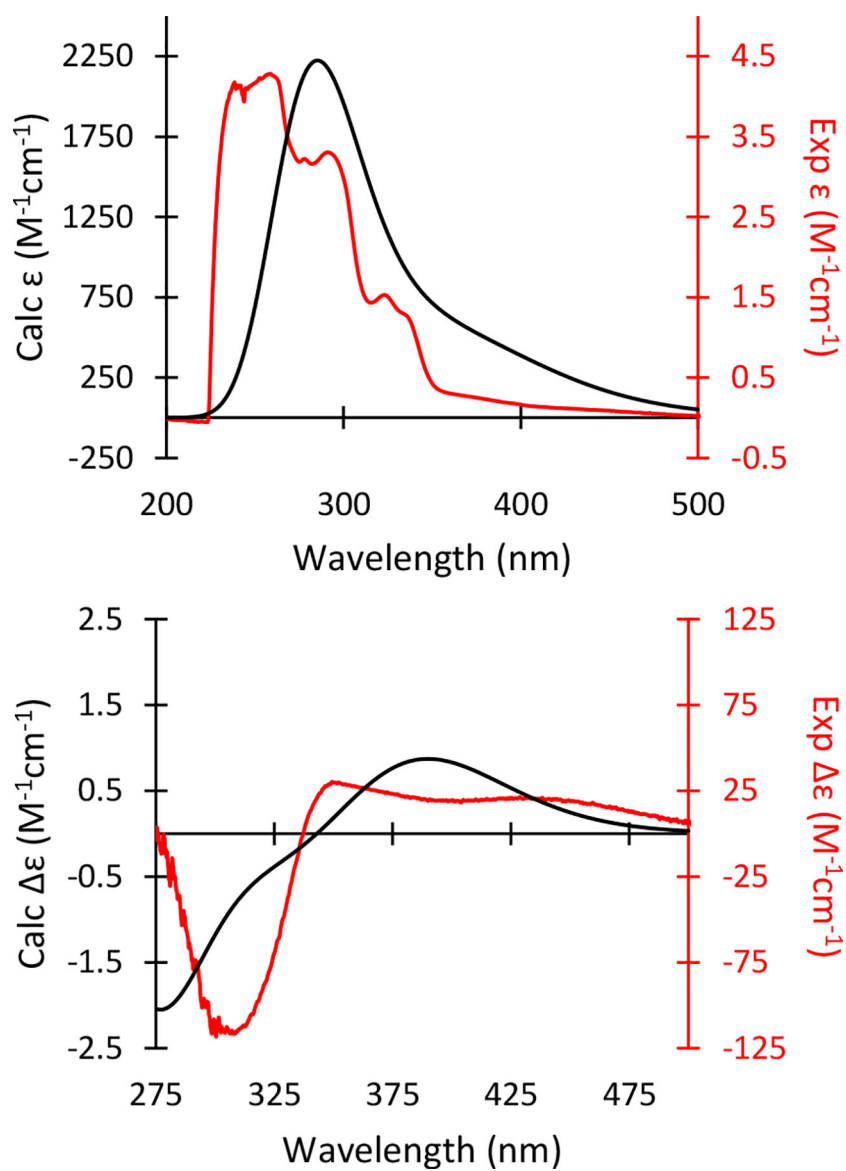
Figure 3:
UV-vis (top) and ECD (bottom) spectra obtained by TDDFT (red) and experiment (black)

Figure 4:
Comparison of predicted and corrected experimental selectivities of alkyl-substituted alkenes. Stereochemistry (*R*/*S*) is indicated by +/− values respectively.
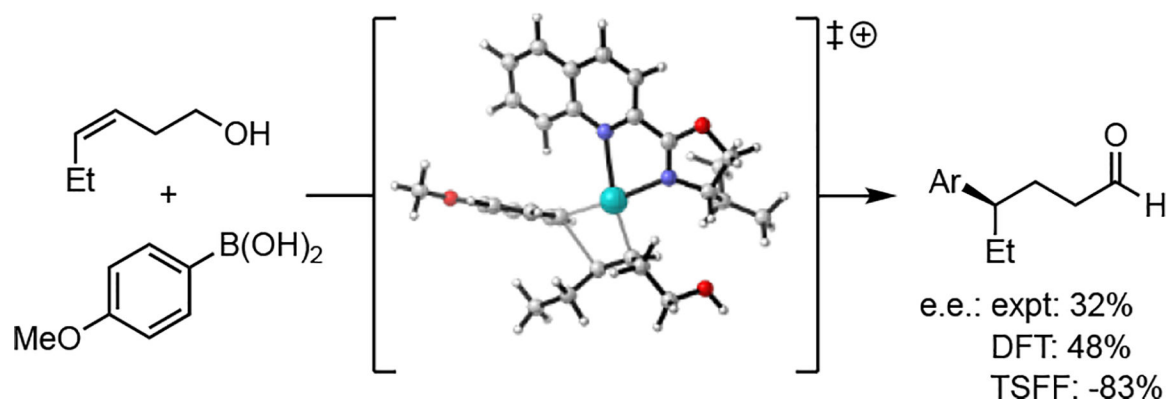
Figure 5:
Ligand screen comparing single-variable correlation and TSFF selectivity predictions. Only the magnitudes of the selectivities are considered. Red lines indicate 2 kJ/mol error.
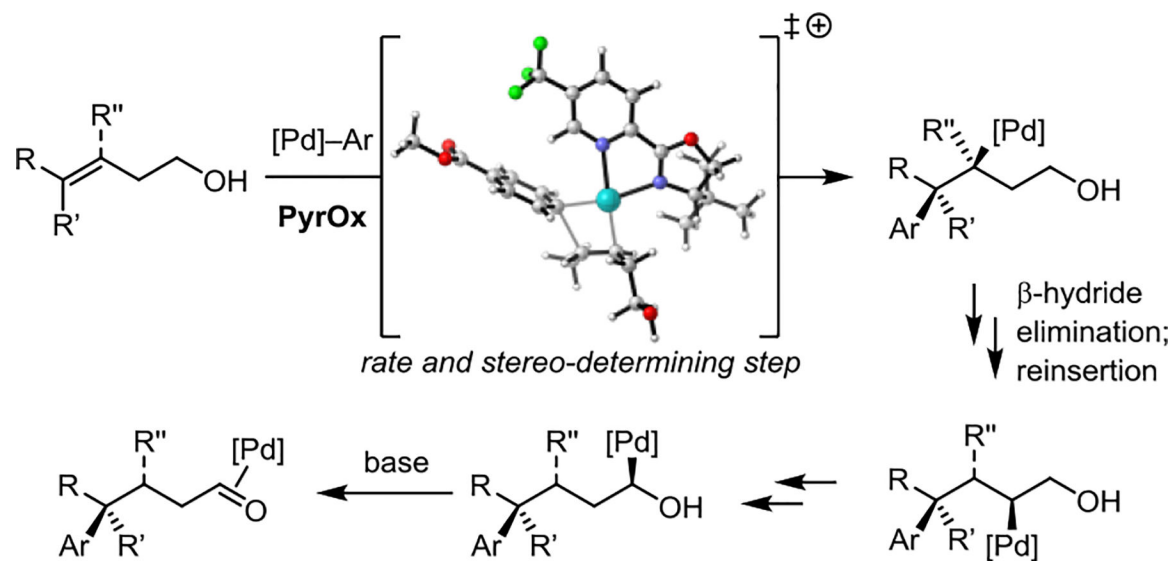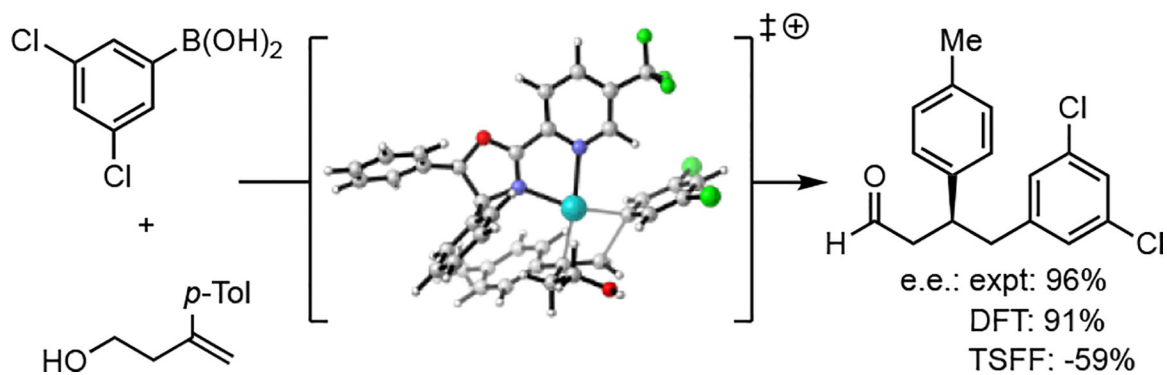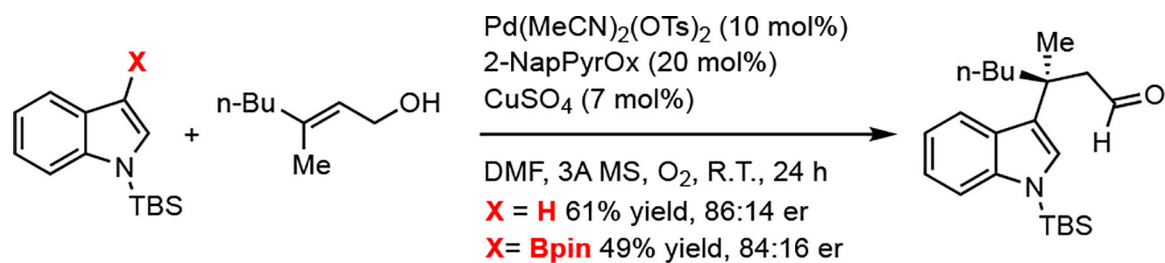
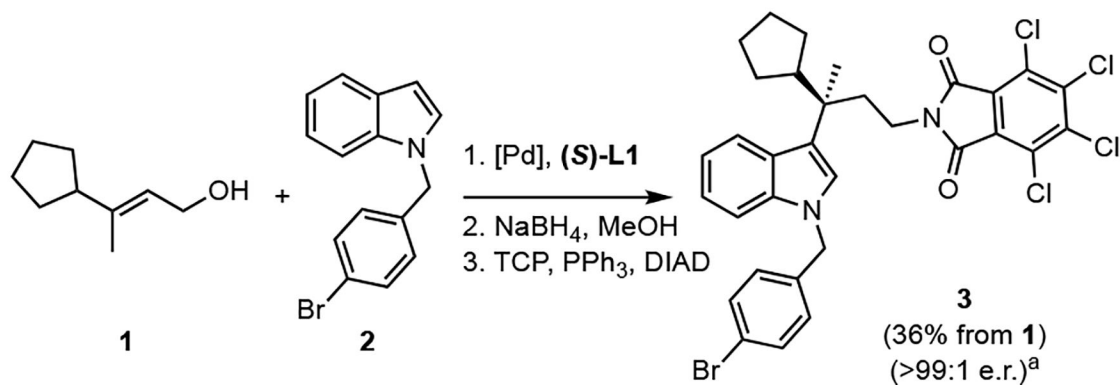**Scheme 1:**
Redox Relay Heck Reaction

Scheme 2:
QuinOx ligands are poorly predicted

Scheme 3:
Arylation of 1,1-disubstituted alkenes

Pd(MeCN)$_2$(OTs)$_2$ (10 mol%)
2-NapPyrOx (20 mol%)
CuSO$_4$ (7 mol%)

DMF, 3A MS, O$_2$, R.T., 24 h
X = H 61% yield, 86:14 er
X= Bpin 49% yield, 84:16 er

b.

1. [Pd], (S)-L1
2. NaBH$_4$, MeOH
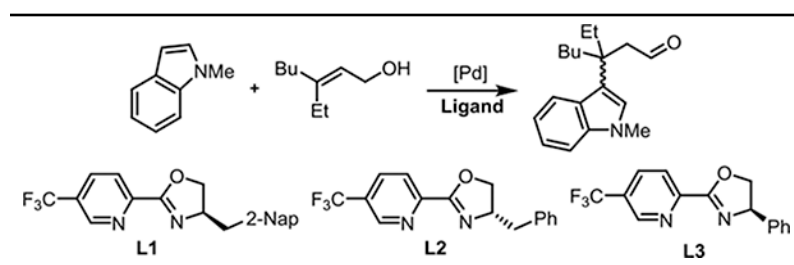3. TCP, PPh$_3$, DIAD

1   2

3
(36% from 1)
(>99:1 e.r.)[a]

Scheme 4:
Experimental verification of major enantiomer[a]. a: e.r. of solution after crystallization

Table 1:

Experimental and Calculated Stereoselectivities

Ligand	Lit	DFT	TSFF
L1	−7.9 (−92%)	12.5 (99%)	4.0 (68%)
L2	−5.7 (−82%)	−6.8 (−88%)	−4.7 (−75%)
L3	3.8 (64%)	9.8 (97%)	2.5 (47%)

Energies are reported in kJ/mol (%ee) where R/S-selectivities are represented as +/−, respectively.