AARON: An Automated Reaction Optimizer for New Catalysts

Yanfei Guan,[†] Victoria M. Ingman,[§] Benjamin J. Rooks,[†] and Steven E. Wheeler[†§*]

[†]*Department of Chemistry, Texas A&M University, College Station, TX 77842*
[§]*Center for Computational Quantum Chemistry, Department of Chemistry, University of Georgia, Athens, GA 30602*
E-mail: swheele2@uga.edu

Abstract

We describe an open-source computational toolkit (AARON: An Automated Reaction Optimizer for New catalysts) that automates the quantum mechanical geometry optimization and characterization of the transition state and intermediate structures required to predict the activities and selectivities of asymmetric catalytic reactions. Modern computational quantum chemistry has emerged as a powerful tool for explaining the selectivity and activity of asymmetric catalysts. However, reliably predicting the stereochemical outcome of realistic reactions often requires the geometry optimization of hundreds of transition state and intermediate structures, which is a tedious process. AARON automates these optimizations through an interface with a popular electronic structure package, accelerating quantum chemical workflows to enable the computational screening of potential catalysts. AARON is built using a collection of object-oriented Perl modules (AaronTools) that provide functionality to build and modify molecular and supramolecular structures. The main functionalities of AaronTools are also available as stand-alone command-line scripts. The core features of AaronTools and AARON are explained and representative applications of AARON to both organocatalyzed and transition-metal catalyzed reactions are presented.

I. Background and Introduction

Despite the widespread success of modern quantum chemistry in explaining the origin of activity and selectivity of homogeneous asymmetric catalytic reactions, the computational design of new catalysts is still far from routine.[1-12] Instead, the vast majority of organic and organometallic catalysts are designed through experimental screening, with applications of quantum chemistry to such reactions done retrospectively. The lack of prospective applications of quantum chemistry to the catalyst design process stems in part from the considerable time and effort required to perform the hundreds of transition state (TS) optimizations required to accurately predict the stereochemical outcome of these reactions.

The selectivity of an asymmetric reaction stems from the difference in relative rates of formation of stereoisomeric products. In simple systems under Curtin-Hammett control,[13] the difference in rates can be attributed to the difference in free energy between stereocontrolling transition states leading to the major and minor stereoisomeric products. However, in most realistic catalytic reactions, there is a breadth of possible mechanistic pathways for a given transformation. Combined with the large conformational space stemming from the flexibility of the substrates and catalyst, the number of accessible TS structures even for a single elementary step in a reaction is often large.[14-15] For practical reasons, quantum mechanical predictions of stereoselectivities are often based on a single low-lying TS structure for the formation of each stereoisomer. However, a growing number of studies on seemingly simple reactions have shown multiple conformations or configurations of a given TS structure, or even different elementary steps, have an impact on selectivity.[12,14] Consequently, making reliable stereochemical predictions requires an exhaustive search for low-lying TS structures to identify all thermodynamically accessible pathways. Manually performing all of these geometry optimizations for different combinations of substrates and catalysts is tedious and time-consuming. As a result, the experimental synthesis and testing of new asymmetric catalysts remains the preferred approach to their development.

Many methods have been developed to automatically locate TS structures connecting two energy minima. Most of these approaches can be summarized as first generating an approximate TS structure, then optimizing this structure to a saddle point on the potential energy surface. From an initial TS structure, optimization algorithms such as the Berny algorithm[16-17] and synchronous transit-guided quasi-Newton (STQN) methods[18] can often locate the exact TS structure. However, the success of these methods depends on the quality of the initial TS

structure provided. Various approaches to generating approximate TS structures have been developed. Single-ended methods systematically adjust the structure starting from a single stable chemical structure, usually the reactant complex, until reaching the TS.[19-26] Double-ended methods require structures of both reactant and product to construct a discretized reaction path along which TS guesses are generated.[27-31] In terms of rapidly localizing many TS structures, a number of automated TS searching and reaction-path exploration approaches have been reported.[32-40] For instance, Morokuma, *et al*.[32-36] introduced the artificial force induced reaction (AFIR) method, which can automatically discover important reaction paths. West *et al*.[37-38] developed a high-throughput automated transition state searching method (AutoTST) for high-throughput computational kinetics. Finally, Jacobsen *et al*.[39] recently published details on the AutoTS code implemented in Jaguar,[41] which generates TS guesses either from a linearly interpolated path connecting reactant and product or by using geometric information from a library of previously computed TS structures.

Despite the power of these methods, the need to sample configurations and conformations, as encountered in asymmetric catalysis, remains an unmet need. For example, in transition-metal catalyzed reactions, there are often multiple configurations of the ligands and substrates around the metal center that lead to thermodynamically accessible TS structures. The multiplicity of relative substrate-catalyst orientations can be even more daunting in organocatalytic systems, particularly those involving ion-pairing.[8,42] At the same time, for flexible substrates or catalysts (as well as those with rotatable substituents, *i.e*. OMe, *i*-Pr, *etc*.), there can be an enormous number of thermodynamically accessible conformations that can cause problems for many of the above-mentioned automated approaches.

There have been many efforts to tackle the problem of molecular conformations, but few of these have focused on sampling transition state conformations. Even fewer are applicable to transition-metal catalyzed reactions. Moreover, many conformational searching approaches rely on classical molecular mechanics (MM) methods, which can be thwarted by the sometimes drastic differences between the QM and MM potential energy surfaces. Consequently, searching for TS conformers is often performed by a filtration strategy, ascending a hierarchy of levels of theories.[8,11,43-45] For example, DiRocco *et al*.[11] developed a workflow that first samples conformations using MM methods, filters and clusters low-lying conformers, then optimizes these using DFT to get a final set of low-lying conformers. Seguin *et al*.[8] employed a similar hierarchical approach combined with extensive manual searching of low-lying conformations.

Because of the absence of appropriate MM parameters for transition metals, semi-empirical methods are typically employed in such cases. For example, Paton *et al*.[44-45] demonstrated the utility of sampling conformations using a Monte Carlo Multiple Minimum (MCMM) algorithm[46] using the semi-empirical method PM6-DH2 followed by DFT optimizations of low-energy conformers. However, Minenkov, *et al*.[47] recently benchmarked the PM6* and PM7 semi-empirical methods for conformations of transition metal complexes, finding that these methods are likely insufficiently accurate to use for screening conformations of such systems. Finally, we note that with all of these approaches, one needs to be able to consistently optimize generated conformations to the nearest TS structure, which can pose technical challenges.

Herein, we describe an open source computational toolkit (AARON: An Automated Reaction Optimizer for New catalysts)[48] that automatically locates multiple conformations and configurations of TS structures for homogeneous catalytic reactions based on user-defined TS templates. AARON does not implement new electronic structure methods or geometry optimization approaches; instead, it automates quantum chemistry application workflows using existing electronic structure packages, thereby alleviating the need for users to construct input files, analyze output files, *etc*. for the hundreds of TS optimizations needed to reliably predict the stereochemical outcome of catalytic reactions. The result is accelerated QM-based predictions of selectivities. AARON is written using a collection of object-oriented Perl modules called AaronTools,[49] which provide functionality for building, manipulating, and comparing molecular structures, constructing input files, parsing output files, analyzing data, and submitting and monitoring jobs using queuing software commonly found on high-performance computer (HPC) clusters. Below, we first summarize the key features of AaronTools and AARON and then present representative applications of AARON to asymmetric transition-metal and organocatalyzed reactions.

II. AaronTools

AaronTools is an open-source collection of object-oriented Perl modules designed to facilitate the construction and analysis of complex molecular structures and the automation of quantum chemistry workflows, with a particular focus on small molecule (homogeneous) catalysts. Similar computational toolkits have been developed by several groups.[50-51,52] For instance, Kulik and coworkers[51] recently introduced MolSimplify, which can rapidly generate reliable structures for transition metal complexes and compute first-principles-based properties.

Central to AaronTools is the `Geometry` class, which provides methods for probing and manipulating molecular structures (distances, dihedral angles, *etc*.) and also serves as a superclass to derive other powerful subclasses. Using the `Geometry` class along with these subclasses, users can create complex molecular and supramolecular structures and prepare input files for computational chemistry packages at various level of theory.[53] Representative molecular structures created using these classes are depicted in Figure 1.

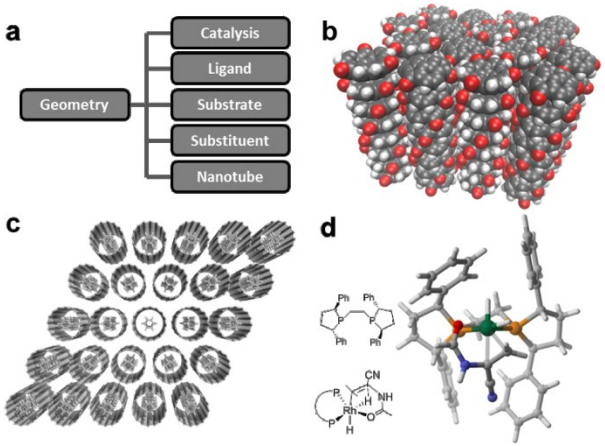


Figure 1. a) Subclasses derived from the `Geometry` class; b) initial solid-state structure of columnar stacks of sumanenetrione;[54] c) solid-state structure of aligned [8,8]-carbon nanotubes with included stacked benzenes; d) transition state structure for a Rh-catalyzed asymmetric hydrogenation reaction.[55]

We will focus on the `Catalysis` and `Ligand` classes below, providing a glimpse of how these tools can be used to construct and manipulate TS structures for asymmetric reactions. The `Catalysis` class contains attributes and methods for either transition state or intermediate structures in catalytic reactions. A `Catalysis` object can be initiated from a standard XYZ file or Gaussian log file. For instance,

```
$cata = new AaronTools::Catalysis( name => 'ts1' );
```

creates a catalysis object based on the geometry read from `ts1.xyz`, which is one of the TS structures reported by Wiest *et al*.[55] for a Rh-catalyzed hydrogenation of enamides using the achiral ligand *Z*-dimethylphosphinoethane (ZDMP, see Figure 2a). Components of the catalysis system (substrates, ligand, and active center—the transition metal in this case) are automatically detected and stored as readily-accessible attributes of this catalysis object.

One can similarly create new ligand objects using the `Ligand` class. AaronTools includes an easily extensible library of common ligands[56] or one can load custom ligands from a user-supplied file. For example,

```
$ligand = new AaronTools::Ligand(name => 'RR-Me-BPE')
```

creates a ligand object containing the chiral ligand (*R*,*R*)-bis(dimethylphospholano)ethane [(*R*,*R*)-Me-BPE]. The backbone and substituents on the ligand are automatically detected and saved as attributes of the object, which facilitates further manipulation of the ligand. Having created this ligand object, one can use `map_ligand` to replace the existing ligand, ZDMP, in ts1 with (*R*,*R*)-Me-BPE (see Figure 2b):

```
$cata->map_ligand($ligand);
```

The resulting structure can be further modified by replacing substituents on the ligand and substrate:

```
$cata->substitute('ligand', 'Me'=>'Ph');
$cata->substitute('substrate', 5=>'COOCH3', 7=>'Me', 8=>'Me');
```

The first `substitute` call converts (*R*,*R*)-Me-BPE into (*R*,*R*)-Ph-BPE by replacing all four methyl groups with phenyl rings, automatically rotating the added substituents to minimize the Lennard-Jones (LJ) energy and then removing any remaining steric clashes by bending and rotating the substituents. The second call of `substitute` modifies the substrate, replacing hydrogens 7 and 8 with methyl groups and the CN group (which starts with atom 5) with $COOCH_3$. These processes are depicted in Figure 2b, and yield a good guess for the corresponding transition state.

Using these and related tools, one can build initial TS structures for virtual libraries of potential catalysts and substrates. For example, initial TS structures for this same Rh-catalyzed hydrogenation reaction can be constructed with the following four lines of Perl for the virtual library of 441 combinations of ligands and substrates depicted in Figure 2c:

```
$lig_subs = [qw(Me Ph tBu iPr CF3 F)];
$sub_subs = [qw(Me Et Ph)];
@cata = $cata->screen_subs('ligand', '24,27'=>$lig_subs, '25,26'=>$lig_subs);
@cata = map {$_->screen_subs('substrate', 7=>$sub_subs, 8=>$sub_subs)} @cata;
```

In particular, atoms (24, 27) and (25, 26) of the (*R*,*R*)-Me-BPE ligand are systematically replaced with all combinations of (H, Me, Ph, *t*-Bu, *i*-Pr, CF_3, and F), while R_1 and R_2 of the substrate are replaced with (Me, Et, and Ph). Similar substitutions can be applied to any molecular structure, rapidly generating diverse libraries of molecular geometries. For instance, in addition to

building libraries of TS structures, these tools can be used to generate structures for the computation of QM-derived molecular descriptors for use in informatics-driven applications.

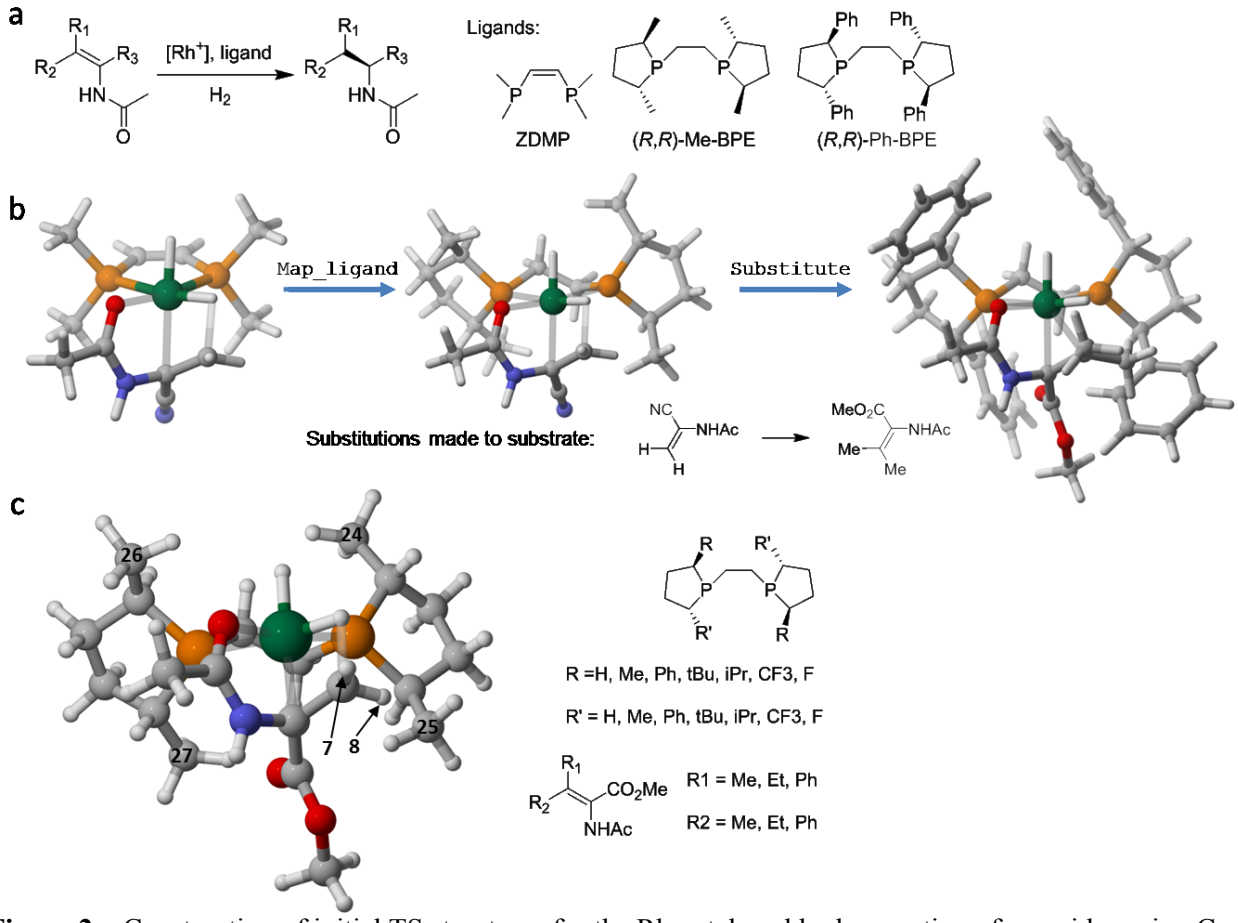


Figure 2. . Construction of initial TS structures for the Rh-catalyzed hydrogenation of enamides using C_2-symmetric phosphoric ligands, starting from a reported TS structure from Wiest *et al*.[55] a) Overall reaction and key ligands; b) replacement of ZDMP with (*R*,*R*)-Me-BPE via `map_ligand` and addition of substituents on both the ligand and substrate via `substitute`. c) Generation of a library of 441 initial TS structures for combinations of 49 ligands and nine substrates.

Finally, we note that these and many of the other components of AaronTools are available as stand-alone command-line scripts. This allows users with no knowledge of Perl to utilize these tools and to incorporate AaronTools functionality into other workflows. For instance, the following system call will replace the ZDMP ligand in `ts1.xyz` with (*R*,*R*)-Me-BPE and write the resulting coordinates to the file `ts2.xyz`:

```
map_ligand ts1.xyz -l RR-Me-BPE -o ts2.xyz
```

III. AARON

Using AaronTools, we have developed a computational toolkit (AARON) capable of automatically and simultaneously screening potential catalysts and substrates for both organocatalytic and transition metal catalyzed reactions based on a user-supplied library of TS and intermediate structures. AARON can rapidly compute the hundreds of TS and intermediate structures required to reliably predict the stereochemical outcome of complex asymmetric catalytic reactions with minimal human intervention, opening the door for the computational screening of potential new catalysts. Briefly, given a previously computed set of TS and intermediate structures for a given reaction, AARON computes analogous structures for related catalysts and substrates while also systematically searching for conformations of rotatable substituents.

A. Overall Program Flow

The overall organization of AARON is shown in Figure 3. It starts by gathering information from a simple and flexible text-based input file. This input includes information about the location of the TS template library and reaction conditions (temperature, solvent, *etc*.) as well as keywords specifying the level of theory (including effective core potentials, custom basis sets, solvent model, DFT integration grid,[57] *etc*.). Preset levels of theory can be defined either system-wide or by each user, allowing for very simple AARON input files for routinely-used levels of theory. In the input file, users also specify catalysts/ligands and substrates to be screened.

AARON constructs each possible catalyst/substrate combination and locates all TS structures for these combinations, following the same workflow for each combination. Briefly, the procedure consists of:

1) building initial TS structures for new ligands and substrates based on the structures in the template library;[58]
2) performing a series of constrained and unconstrained geometry optimizations (Steps 1-3), followed by harmonic vibrational frequencies (Step 4) and potential higher-level single point energies (Step 5); and
3) analyzing the resulting structures.

There are typically many TS structures for each substrate/catalyst combination, and Steps 1-5 for each catalyst/substrate combination are run as separate jobs in a typical HPC environment. This

results in high throughput, particularly for HPC resources with a large number of individual nodes.

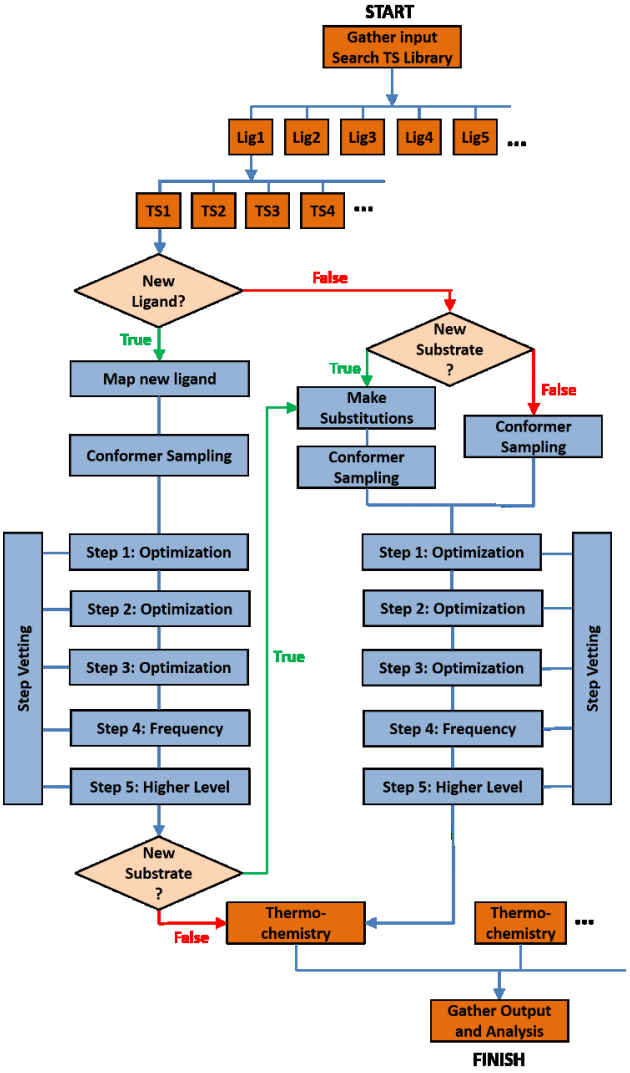


Figure 3. Overall workflow of AARON.

To maximize efficiency when varying both catalyst and substrates, AARON performs a hierarchical search in which TS structures are first located for any new ligands acting on the original substrates found in the template library (the left part in Figure3). Once these TS structures have been located and vetted, they are used as templates for any new substrates (the right part in Figure 3). If neither new ligands nor substrates are requested, AARON performs a conformational search of rotatable substituents based on the structures in the TS template library.

After AARON builds initial TS structures, it performs a series of constrained and unconstrained optimizations (Steps 1-3) starting from the initial structure built by AaronTools. In

Step 1, any new component of the structure (*e.g.* a new ligand or substituent) is optimized at a low level of theory (the default is PM6). This refines the initial TS structure and removes steric clashes, reducing the number of steps in downstream optimizations at higher levels of theory. In Step 2, an energy minimization is performed at a user-defined level of theory with constraints added to all forming/breaking bonds to yield a structure that should closely resemble the transition state. In Step 3, AARON performs a full TS optimization using the Berny algorithm built into Gaussian.[16-17] At each point in this process there are many possible issues that can arise; these are caught by the TS vetting process (see Section F below). After a possible TS structure has been located, harmonic vibrational frequencies are computed to confirm the nature of the stationary point and to obtain thermochemical data. If requested, single point energies are also computed at a higher level of theory.

AARON employs a flexible and modular workflow that can be readily deployed on HPC clusters with popular queueing systems and can be stopped and restarted at any point. Input and output files for the electronic structure computations are stored and organized under an easily-navigated directory tree that allows users to monitor the workflow and fix any problematic structures encountered. A utility is included that automatically constructs tables of absolute energies, enthalpies, and free energies as well as all unique optimized structures for inclusion in Supporting Information. Capabilities are currently in development that will populate databases with computed TS properties for informatics applications.

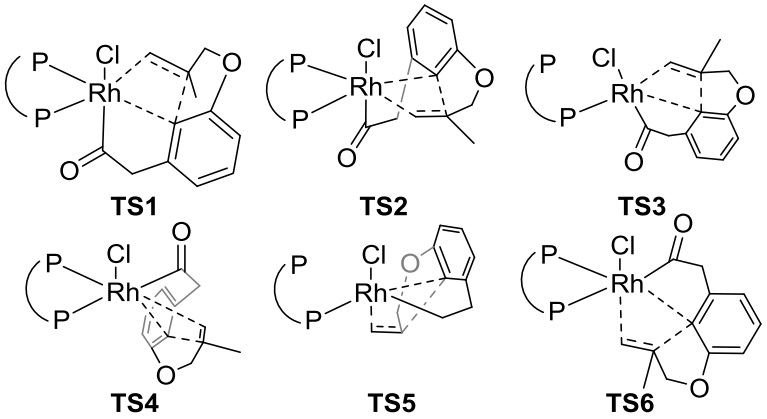
B. Construction of Initial Transition State Structures

Central to locating TS structures is the generation of an initial TS guess for a given catalyst and substrate. AARON uses geometries from previously located TS structures (the TS template library) as templates for the new catalysts and substrates. This approach is also one of the core strategies used to generate initial TS structures in AutoTS.[39] In contrast to AutoTS, however, which automatically detects the templating TS using SMILES strings, AARON requires the user to specify the reaction template. While less general, such an approach allows AARON to handle complex catalytic reactions that are still out of reach of more fully automated approaches. The TS template library is stored as standard XYZ files in a plain-file database. In addition to a system-wide TS template library, users can construct their own libraries.

AARON will compute analogous structures for new catalysts/substrates for all structures found in the specified library, automatically searching over conformations of rotatable

substituents. Thus, AARON partially addresses the configuration problem by having the user enumerate the TS configurational space for a simplified ligand/substrate manually by including these configurations in the TS template library. Consequently, the library should contain all reasonable configurations and relative substrate-catalyst orientations that can lead to thermodynamically accessible TS structures for any of the catalysts and substrates to be screened. This can consist of TS structures for multiple elementary steps as well as the corresponding intermediates—whatever is needed to predict the desired selectivity (although the focus here is on stereoselectivities, AARON is also suitable for predicting regioselectivites, *etc*.).

An example of a TS template library can be taken from work from Liu *et al*.,[59] who conducted a thorough study of the regio- and stereoselectivity of Rh-catalyzed carboacylations of benzocyclobuteneones. Liu *et al*.[59] considered six configurations of the stereocontrolling olefin migratory insertion step for this reaction (**TS1**-**6** in Scheme 1), optimizing these structures for the achiral ligand dppb to identify the most favorable configuration (TS1). Then, the diastereomeric forms of TS1 leading to the two possible stereoisomeric products were computed to understand the stereoselectivity of the chiral ligand SEGPHOS. However, there is the possibility that some of the other TS structures in Scheme 1 could impact the selectivity for the SEGPHOS-catalyzed process even though they are relatively high-lying for dppb. Using the dppb-based structures already computed by Liu *et al*.[59] as a TS template library, AARON could readily compute all structures leading to both the major and minor stereoisomer for the chiral ligand SEGPHOS.



Scheme 1. Example of the multiple configurations and substrate-catalyst orientations that must be included in a TS template library for a Rh-catalyzed carboacylation reaction studied by Liu *et al*.[59]

Once the TS template library has been constructed, AARON can be used to make predictions for different ligands or catalysts (for simplicity, we will primarily refer to ligands below). AARON maps the donor atoms of any new ligand to the corresponding atoms of the ligand

found in the TS template library (for transition-metal based catalysts, ligand atoms bound to the metal are automatically detected as donor atoms; for organocatalysts, the donor atoms must be identified when constructing the TS library). The only information required are the identities of the donor atoms for the new ligand (*i.e.* those that either bind the transition metal or, for organocatalysts, engage directly with the substrates). These donor atoms are mapped onto the analogous atoms in the templating TS to mimic the template structure as closely as possible. The precise mapping strategy depends on the nature of the ligand/catalyst, which is detected automatically. AARON can map a broad range of ligands/catalysts for both organocatalyzed reactions and transition metal catalysts, including multidentate, bidentate, and monodentate ligands (see Figure 4) as well as 'multi-block' systems with donor atoms located on different catalyst components connected via flexible covalent linkers (see Figure 5).

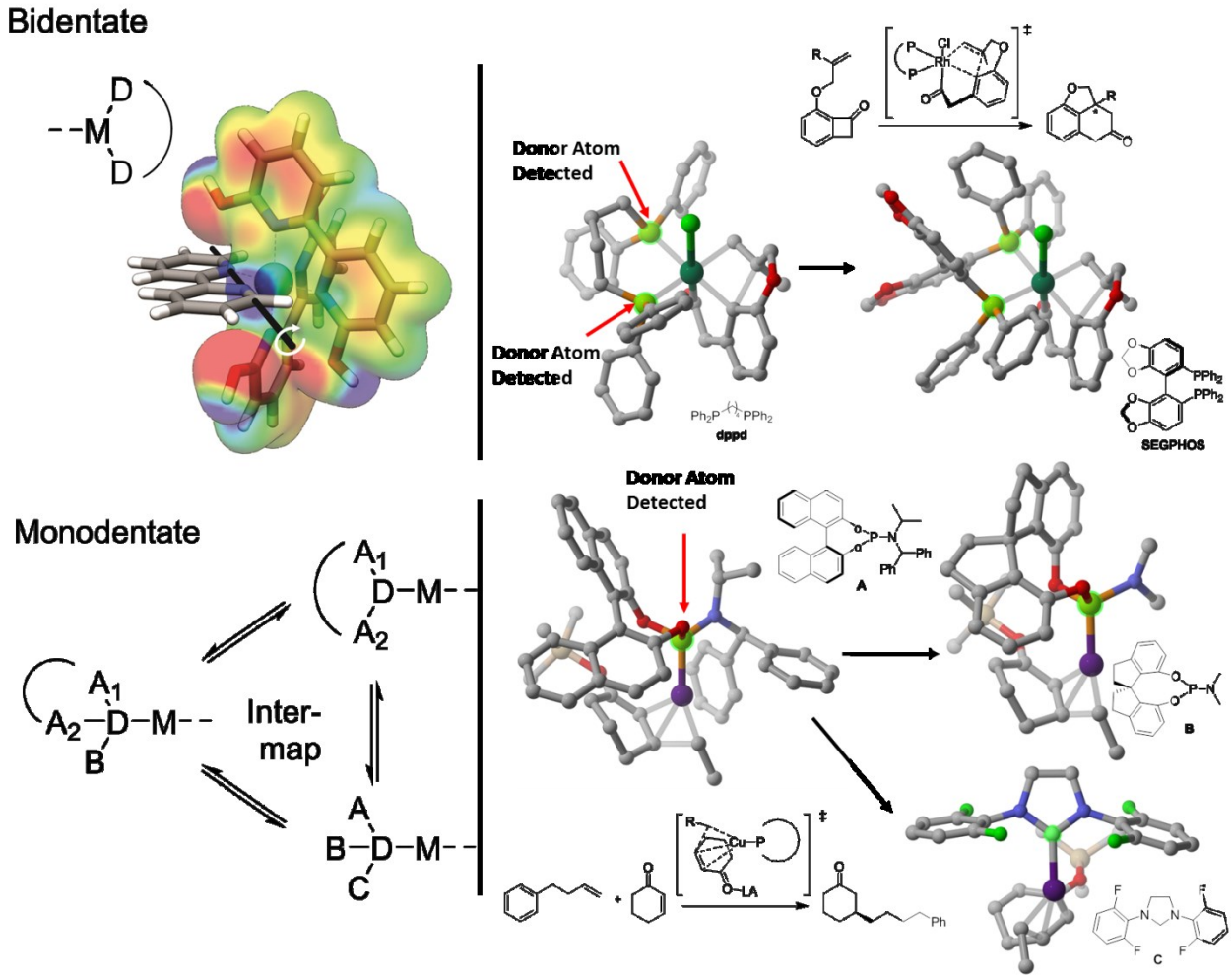


Figure 4. Examples of replacing bidendate and monodentate ligands. Donor atoms are highlighted in green.

For multidentate ligands, AARON maps new ligands by minimizing the root mean-squared deviation (RMSD, determined using the quaternion-based algorithm of Kearsley[60]) between all donor atoms of the new and old ligands. For bidentate ligands, the two donor atoms alone do not uniquely determine the ligand position. Therefore, after positioning the ligand to minimize the RMSD for the two donor atoms, the ligand is rotated around the axis defined by these donor atoms to minimize the LJ energy of the system. For example, Figure 4 shows the replacement of the bidentate ligand dppb with SEGPHOS in TS1 from Scheme 1.[59] For monodentate ligands, the ligand can fall into several classes depending on the presence of cyclic substructures and AARON can automatically map among these different classes. For example, Figure 4 shows the replacement of the BINOL-based phosphoramidite ligand **A** in a TS structure for a Cu-catalyzed conjugate addition computed by Paton, *et al.*[44] with both a SPINOL-based analog (**B**) and an NHC (**C**).

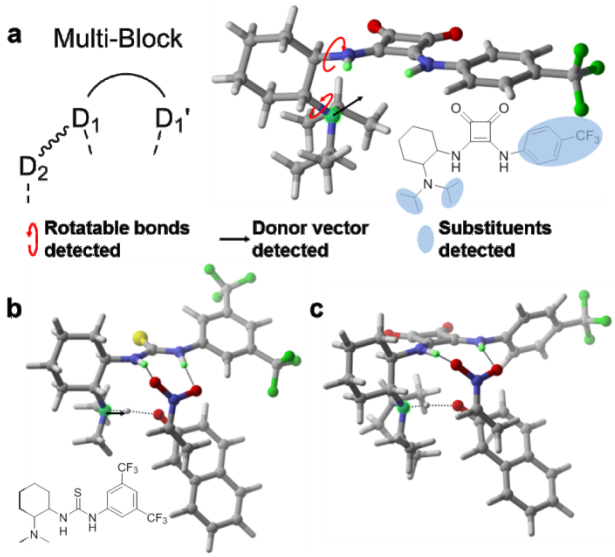


Figure 5. Example of replacing a 'multi-block' catalyst in which multiple atoms of the catalyst connected by flexible linkers bind the substrate. a) thiourea-catalyzed addition of nitroethene to a β-napthol reaction and the key TS. b) Mapping of a squaramide-based catalyst onto the TS structure for a thiourea-based catalyst.

More general binding motifs ('multi-block' binding) occur for many organocatalyzed reactions, in which internal torsions can change the relative position of catalyst atoms that bind the substrate (still called 'donor atoms' here for simplicity). AARON first maps donor atoms connected to the core of the catalyst ('central donors') and then those connected to peripheral groups ('remote donors'). The central donors are mapped using the above strategy for

monodentate, bidentate, *etc*. ligands depending on the number of donor atoms. Rotatable bonds connecting the central donors and remote donors are detected automatically and the dihedral angles around these bonds sampled to best overlap the donor vector between remote donors of new catalyst and those found in the TS template library while also adjusting the position of the central component (if applicable) of the catalyst to minimize steric crowding. Substituent orientations and angles are then altered to remove remaining steric clashes to generate a reliable initial TS structure. For example, for the asymmetric dearomatization reaction catalyzed by a bifunctional hydrogen bonding catalyst presented in Figure 5a,[61] a thiourea-based organocatalyst is replaced by a squaramide-derived catalyst with a different backbone and sterically-congested tertiary amine (Figure 5b).

D. Conformational Searches

While the configurational space must be manually enumerated in the construction of the TS template library, AARON can automatically search portions of the conformational space spanned by rotatable substituents.[62] This represents a considerable time-savings for the user, as exhaustively searching these conformations manually is tedious. AARON utilizes a rule-based hierarchical search strategy based on knowledge of preferred torsional angles for each substituent type to prevent the combinatorial explosion of the conformational space that would result from a more brute force approach.[63] The rules are: 1) conformers are sampled only for substituents detected by AARON or specified by the user as new substituents; 2) torsional angles sampled for each substituent are determined based on its symmetry; 3) conformations are searched hierarchically such that conformational searching for later substituents begins only after completion of the conformer search for earlier substituents. In contrast to many of the tiered conformational search algorithms[8,11,43-44] that rely on initial sampling with MM-potentials or semi-empirical methods, AARON searches conformations directly at the chosen DFT level. This circumvents issues arising from differences between MM or semi-empirical and DFT potential energy surfaces.

The first rule accounts for the fact that many conformational changes are inconsistent with a given TS structure. For instance, potentially flexible linkers connecting donor atoms are recognized as the backbone of the ligand and are automatically rotated to position the donor atoms to best bind the metal center (in transition-metal catalysis) or substrates (in organocatalysis). The second rule relies on the fact that generated conformers provide initial TS

structures that will then be optimized. As such, there is little need for dense conformational sampling since most of these TS guesses will converge to identical structures. The conformational space to be explored is reduced by sampling conformations according to the symmetry of substituents. Despite this, for catalysts or substrates with even a modest number of rotatable groups (> 5) there can be an enormous number of combinations of conformers to be sampled. This necessitates the hierarchical searching method, as per rule 3. In this hierarchical searching method (see Figure 6), AARON first searches conformers for one substituent. Starting from all unique conformers identified by rotating this substituent, AARON generates a new generation of conformers by rotating the next substituent, and so on. The result is the hierarchy of conformers depicted in Figure 6.

Even though these rules lead to sampled structures that are sparsely distributed in the conformational space, many initial TS structures converge to the same conformers. To account for this, AARON monitors all conformers during the geometry optimizations. If a duplicate conformation is identified (based on an RMSD cutoff of 0.15 Å), the corresponding job is killed and the repeated conformer removed. As a result, children conformers of the repeated conformer are never sampled. While this hierarchical check-and-remove mechanism is intended to accelerate the conformational search, it may miss key conformations, especially in cases in which substituents engage in any sort of 'geared' conformational change (*e.g.* multiple *t*-Bu groups on nearby carbons). For these cases, a full, brute-force conformer search can be performed.

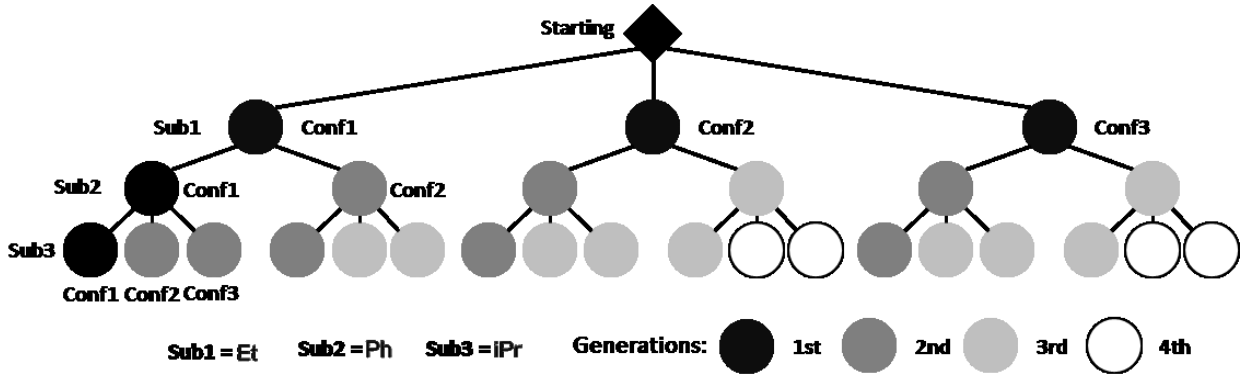


Figure 6. Example of the hierarchical searching of a fictitious system with three substituents (Et, Ph, and ^{i}Pr).

E. Thermochemistry

AARON provides predicted selectivities based on computed energies, enthalpies, and free energies. The latter are computed using both the standard rigid-rotor/harmonic-oscillator

approximation (RRHO) and the quasi-RRHO approximation of Grimme.[64] These values can be computed based on the level of theory used for the geometry optimizations and vibrational frequencies or using higher-level single point energies. In the latter case, the entropic contributions are computed using the vibrational frequencies from the lower level of theory.

After the conformational search, there are typically many unique conformers corresponding to each TS structure from the TS template library. For each thermodynamic quantity (energy, enthalpy, free energy), AARON performs a Boltzmann weighting over all unique conformers to compute the effective energy for each TS. For instance,

$$G_{effective} = -RTln\left(\sum_{i}^{conformers} e^{-\frac{G_i}{RT}}\right)$$

Final predicted stereoselectivities are based on a Boltzmann-weighted sum over the effective energies, enthalpies, and free energies for all TS structures leading to formation of different stereoisomers. For example, for an enantioselective reaction the *ee* is computed in terms of effective free energies as

$$ee(\%) = \frac{\sum_i^{TS} e^{\frac{-\Delta G_{eff}(R_i)}{RT}} - \sum_i^{TS} e^{\frac{-\Delta G_{eff}(S_i)}{RT}}}{\sum_i^{TS} e^{\frac{-\Delta G_{eff}(R_i)}{RT}} + \sum_i^{TS} e^{\frac{-\Delta G_{eff}(S_i)}{RT}}}$$

F. Step Vetting and Error Checking

Since transition state optimizations are prone to failure, AARON checks all running jobs periodically to ensure that they are converging to the correct TS structure ("Step vetting" in Figure 3). If a job aborts for any reason or takes an incorrect geometry optimization step (*e.g.* breaks or forms a bond not involved in the targeted elementary step), AARON will attempt to either fix the structure or add additional keywords to the input file to address the problem (Figure 7). AARON compares the distances for any forming/breaking bonds in the transition state structure with those of the templating TS. If the distance between reacting atoms is too large, AARON will shorten this coordinate and restart from Step 2, and vice versa. AARON also monitors the connectivity of the system during all optimizations. If any unexpected bond breaking or forming is identified, AARON stops the current optimization, fixes the geometry, and restarts from Step 2 with additional constraints added for this problematic bond coordinate. Upon passing TS vetting, AARON checks the output file for any errors and responds accordingly based on a set of rules derived from our experience locating TS structures. This includes errors

arising from SCF convergence failure, *etc*. AARON tracks the number of attempts for each step, and adds additional keywords (for example, to shorten the maximum step size or compute force constants more frequently) to try to deal with particularly problematic optimizations. The net result is a robust protocol to consistently locate TS structures with precise error control and geometry vetting.

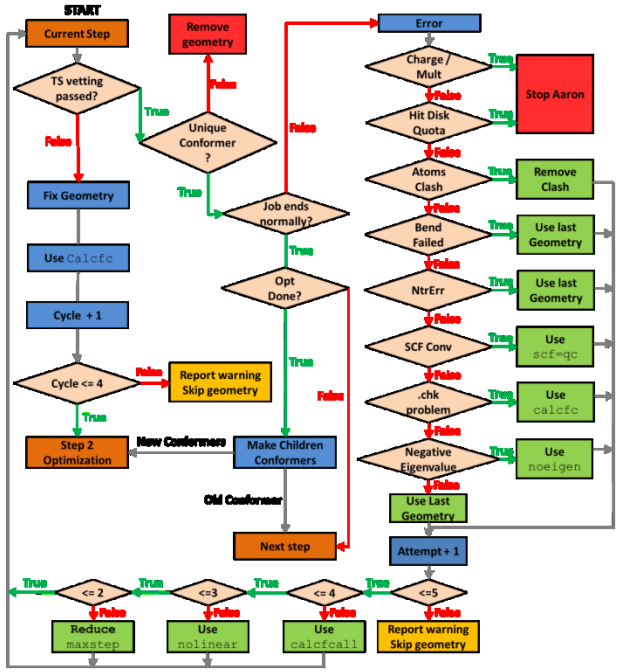


Figure 7. Procedure for 'Step Vetting', in which AARON constantly monitors all jobs to ensure that correct TS structures are located. For jobs that die with no recognizable error, the last geometry is used and the job restarted.

IV. Representative Applications

Below, we demonstrate the power of AARON through applications to various reaction types. We strove to design AARON to be as flexible and general as possible, and the following examples show that AARON can be applied broadly.

A. Pd-Catalyzed Heck Allenylation

Sigman and co-workers[65-68] have developed a series of Pd-catalyzed redox-relay Heck reactions and Wiest, Sigman, and co-workers[69] have shown computationally that the migratory insertion step in these reactions is stereocontrolling. Most recently, Sigman *et al*.[70] studied the Heck relay reactions of substrates **2a** and **2aa** using the chiral ligand 2-t-Bu-PyrOx (**L1** in Figure

8a) at the M06/6-31+G(d)/LANDL2DZ level of theory. They performed a systematic search of the four possible orientations of the substrates relative to the catalyst (TS1 – TS4 in Figure 8b) for formation of the two enantiomers of the product. For some of these, they also considered multiple conformations of the hydroxymethyl group. In total, thirteen low-lying TS structures were reported for the reaction of **2a** and nine for **2aa**. The experimental selectivity, which is significantly greater for substrate **2a** than for **2aa**, was explained in terms of the difference in free energy between the lowest-lying TS structures leading to the two different enantiomeric products.

This reaction provides an ideal system to demonstrate the ability of AARON to locate TS structures covering the full conformational space of the catalyst and substrates. We constructed a TS template library using the eight unique TS structures (ignoring the conformations of the CH_2OH group) reported by Sigman *et al*.[70] The cyclohexenone of substrate **1** can potentially have two ring-puckered conformations, which doubles the number of structures that must be included in the TS template library to 16. We used AARON at the same level of theory employed by Sigman *et al*.[70] to locate all low-lying TS structures for substrates **2a** and **2aa** using this TS template library and systematically searching for conformations of the substituents CF_3, *t*-Bu, and CH_2OH. Systematically considering all orientations of the substituents combined with the two conformations of the cyclohexenone leads to 192 possible TS structures for each substrate (**2a** and **2aa**). Without any user intervention, AARON located 90 unique TS structures for **2a** and 94 for **2aa**. The 'missing' TS structures are primarily due to conformations that converged to identical structures during the optimizations.

Relative free energies for the computed TS structures for substrates **2a** and **2aa** are plotted in Figure 8c. Overall, we find that there is a dense manifold of thermodynamically accessible TS structures lying within 2.6 kcal/mol of the global minimum energy TS structure (white and light-gray shaded regions in Figure 8c). For example, for substrate **2a**, 21 of the 90 computed TS structures are in this range, while for **2aa** this number increases to 49 of 94. In addition to showcasing the ubiquity of low-lying TS structures for this seemingly simple transformation, the data in Figure 8c indicates that there is not a single key low-lying (*R*) and (*S*) transition state structure for either substrate. Instead, multiple TS structures arising from different TS configurations impact the stereoselectivity. Moreover, the configuration giving rise to the lowest-lying TS structure for each stereoisomer differs for the two substrates. These data highlight the

challenges of improving the selectivity of a given catalyst—it is often not sufficient to simply increase the free energy difference between a single pair of TS structures leading to different stereoisomers; instead, one must devise a catalyst that differentiates between two disparate families of TS structures. These results also highlight the dangers of manually searching for stereodetermining transition states using conventional tools, since one can easily miss key low-lying structures. The lowest-lying TS for the formation of each stereoisomer can arise from qualitatively different configurations for different substrates, and assumptions regarding the 'preferred' configuration based on results for one substrate are often not transferrable to other, seemingly similar substrates.

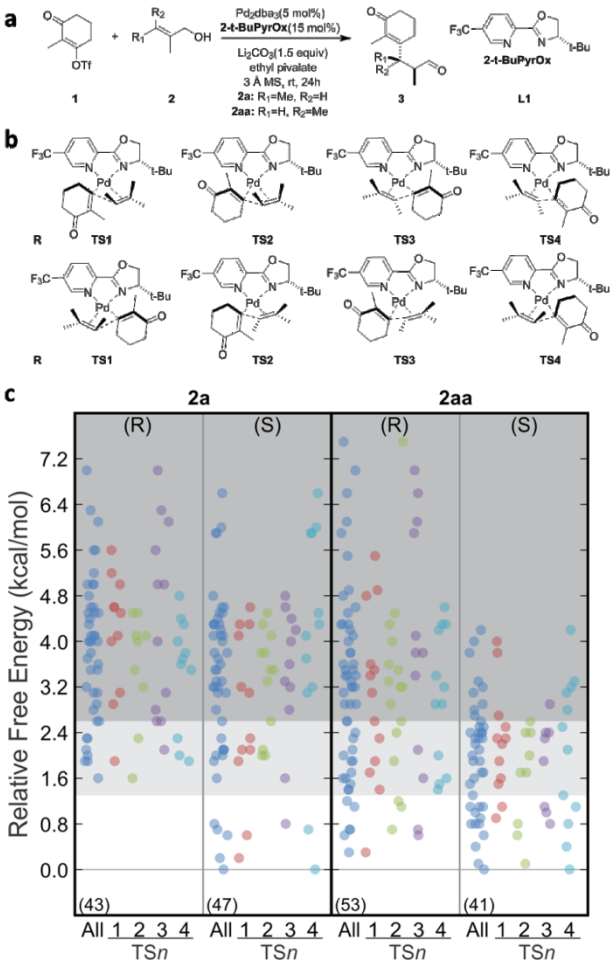


Figure 8. a) Pd-catalyzed redox-relay Heck reaction from Sigman and co-workers[70] using the chiral ligand 2-t-BuPyrOx. b) Eight TS structure comprising the TS template library for the application of AARON to this reaction. c) Relative free energies for TS structures located by AARON for substrates **2a** and **2aa**. The gray shaded regions denote energy ranges of TS structures expected to have Boltzmann populations of <1% (dark gray) and <10% (light gray) relative to the lowest-lying TS structures.

A key benefit provided by AARON is that studying two substrates requires no more user effort than studying one, in contrast to the manual application of conventional computational tools. Indeed, one can screen a given catalyst across dozens of substrates by simply specifying the substitutions that must be performed to construct this substrate library. In the following examples, we will discuss the use of AARON to screen multiple catalysts for a given reaction.

B. Rh-Catalyzed Hydrogenation of Enamides

Transition-metal-catalyzed asymmetric hydrogenations represent a cornerstone of synthetic routes to a wide range of optically active compounds,[71] and chiral phosphorus ligands have proved highly effective in catalyzing this class of reactions. However, some mechanistic details of Rh-catalyzed hydrogenations of enamides are still unsettled. AARON was previously used to probe the mechanism and to explain the stereoselectivities of the Rh-catalyzed asymmetric hydrogenation of enamides shown in Figure 9.[12] Previous computational studies[72-75] and our preliminary transition state searches revealed two key steps for this reaction in which H is transfered from the Rh to the alkene group of the substrate (Steps 1 and 2). Two distinct mechanisms are possible for this reaction depending on whether the α- or β- carbon is hydrogenated first. In each step for each mechanism, the substrate can adopt different configurations with respect to the Rh atom, giving rise to at least eight potential TS structures leading to each of two enantiomeric products. This, combined with the existence of rotatable OMe groups on both the substrate and some of the ligands, leads to hundreds of possible TS structures for each ligand.

Using AARON, we screened six ligands and located a total of ~250 TS structures spanning the configurational and conformational spaces. Unlike the previous example discussed, in which there was a single stereodetermining step but multiple configurations and conformations, this reaction is more complex in that there are multiple mechanisms and stereodetermining steps combined with different configurations and conformations. This leads to numerous reaction paths. Figure 9 shows the relative free energies of TS structures for **L2**, for which the predicted stereoselectivity is in exact agreement with experiment.[76] There are paths connecting many of these TS structures for the two hydrogenation steps and the two possible mechanisms, leading to a complex web of possible reaction paths. For each reaction path, the step with higher free energy is rate-limiting. As seen in Figure 9, for some pathways the first H-transfer is rate-limiting whereas the second step is rate-limiting for others. Furthermore, there are a large

number of thermodynamically accessible pathways and both stereoisomers can form via multiple mechanisms. Indeed, the lowest-lying pathway to (*S*) is predicted to lie below the lowest-lying pathway leading to (*R*); predictions of the stereoselectivity based solely on these two low-lying TS structures would incorrectly predict the overall sense of stereoinduction. However, a Boltzmann weighting over all low-lying pathways correctly predicts the excess formation of the (*R*) product due to the existence of several relatively low-lying pathways leading to this stereoisomer. Thus, there is not a single step or even mechanism that completely characterizes this reaction, and reliable predictions of stereoselectivity require a Boltzmann weighting over all accessible pathways. AARON performs this weighting automatically. Moreover, the data provided by AARON showed that the mechanism corresponding to the lowest-energy path for this reaction varies with the structure of the ligand, highlighting the importance of considering all viable reaction paths for each ligand screened.

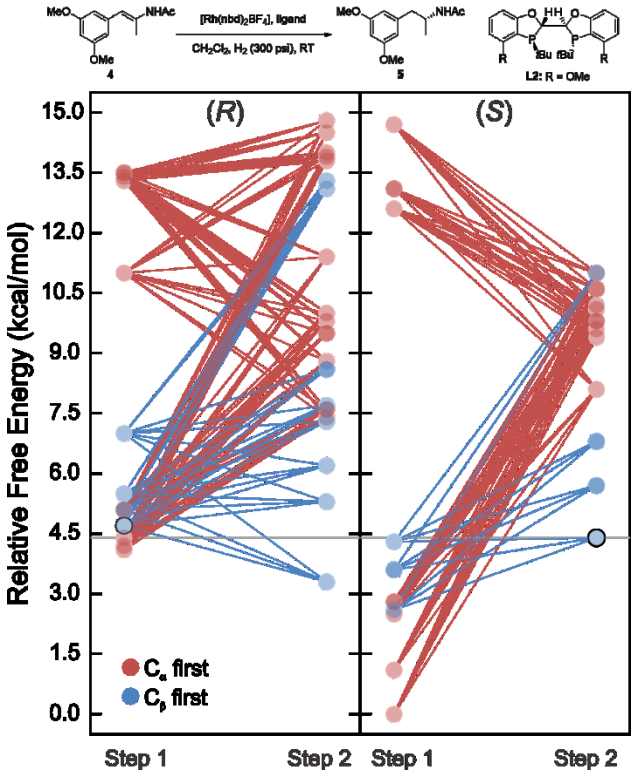


Figure 9. Relative free energies for TS structures for the asymmetric hydrogenation of enamides for the first and second hydride transfer steps (Steps 1 and 2) following two possible mechanisms (hydride transfer to C_α first or C_β first) for formation of the major (*R*, left) and minor (*S*, right) stereoisomeric products. Colored lines indicate pathways between the TS structures for two steps. For formation of each stereoisomer, the lowest-lying rate-limiting TS structure is outlined in black. Data from Ref. [12].

C. Lewis-Base Promoted Propargylation of Aromatic Aldehydes

Finally, we note that AARON has been successfully applied to organocatalyzed reactions, more specifically bidentate Lewis-base catalyzed allylations and propargylations of aromatic aldehydes.[77-78] The stereocontrolling step in these reactions involves a hexacoordinate silicon, and previous computational studies established that there are five distinct ways to arrange the six groups surrounding the Si.[79] Each of these five configurations results in a pair of TS structures leading to each of the possible enantiomeric alcohols, leading to ten potential TS structures for C_2 symmetric catalysts and 20 for non-C_2-symmetric ones.[14] Lu and co-workers[14,80] showed that accurate stereochemical predictions for these reactions require the computation of all ten or 20 possible structures.

Starting from a TS template library constructed from the TS structures of Lu *et al*.,[80] Rooks *et al*.[77] used AARON to screen a set of 18 bipyridine *N*,*N*′-dioxide derived catalysts for the allylation of benzaldehyde for which experimental stereoselectivities were available. The computed stereoselectivities were in good agreement with experiment, with predictions for 16 out of 18 catalysts within 20% of the experimental data. AARON was also used to screen these same catalysts for the asymmetric propargylation of benzaldehyde, revealing several catalysts predicted to be moderately stereoselective. More recently, Doney *et al*.[78] used AARON to screen a library of 60 potential catalysts built on six bipyridine *N*,*N*′-dioxide-derived scaffolds for this same propargylation reaction. Predicted *ee*'s ranged from 45% (*S*) to 99% (*R*), with 12 catalysts predicted to exhibit stereoselectivities exceeding 95%. The analysis of a large number of catalysts also revealed broad trends in the origin of stereoselectivity in this reaction that would have been difficult to unravel by studying a more limited number of instances. Finally, based on these data, Doney *et al.*[78] proposed a novel catalyst predicted by AARON to provide very high stereoselectivities by preferentially stabilizing a particular TS structure.

V. Summary and Concluding Remarks

The ability to reliably and rapidly predict the stereoselectivities of asymmetric catalytic reactions across different chiral ligands and substrates is a prerequisite for computational catalyst design.[1-3] We have described an open-source computational toolkit (AARON) that can 1) automatically generate initial TS structures for new ligands and substrates based on a library of TS templates, 2) optimize TS structures with precise error control and geometry vetting, 3)

search conformers in a parallel and hierarchical way, 4) screen combinations of ligands and substrates from a simple input file, and 5) predict selectivities through a Boltzmann weighting of multiple TS structures or reaction paths. This is accomplished with the aid of a collection of object-oriented Perl modules (AaronTools) designed to facilitate applications of quantum chemistry to complex molecules. AARON also includes a number of utilities to facilitate the organization, storage, and dissemination of generated data.

We have demonstrated that AARON can be applied to both organocatalysis and transition metal catalysis, locating far more low-lying TS structures than can reasonably be found manually. The importance of considering multiple configurations and conformations across multiple elementary steps to reliably predict stereoselectivities was highlighted for several asymmetric reactions. The sheer number of TS structures that must be computed in such cases is best handled with automated tools such as AARON, and studies relying on the manual application of quantum chemical tools often neglect such subtleties.

There is ample room for improvement of AARON, which is still in early stages of development. The most glaring need is an automated approach for constructing the TS template libraries on which AARON relies.[81] There have been a number of heuristic-guided configuration searching methods developed that are applicable to catalytic reactions,[21,82-87] and methods based on bond-connectivity have been used to explore multiple reaction pathways for transition metal catalyzed reactions.[83,65,67] We anticipate combining AARON with such tools in order construct a more comprehensive package for computational catalyst design.

Finally, we note that the large quantities of structural and energetic transition state data generated by AARON are ripe for the application of modern informatics tools. For instance, there has been success in applications of multivariate regressions to experimentally-generated stereoselectivity data as a means of both understanding stereoselectivity and designing better catalysts.[88] AARON opens up the door to such informatics-driven catalyst design efforts based solely on computed data. Such studies are currently underway.

Acknowledgments

This work was supported by the National Science Foundation (Grant CHE-1665407). We also thank Tongxiang "Aaron" Lu, E. C. Sherer, M. R. Haas, A. N. Bootsma, S. Gallarati, and D. P. Harding for valuable discussions. Portions of this research were conducted with high

performance research computing resources provided by Texas A&M University (http://hprc.tamu.edu). All electronic structure computations were performed using Gaussian09[89] and all molecular structure figures were generated using CYLView.[64]

Supporting Information Available: Cartesian coordinates, absolute energies.

References

1. Houk, K. N.; Cheong, P. H.-Y., Computational Prediction of Small-Molecule Catalysts. *Nature* **2008,** *455*, 309.
2. Houk, K. N.; Liu, F., Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017,** *50*, 539-543.
3. Poree, C.; Schoenebeck, F., A Holy Grail in Chemistry: Computational Catalyst Design: Feasible or Fiction? *Acc. Chem. Res.* **2017,** *50*, 605-608.
4. Hopmann, K. H., Quantum Chemical Studies of Asymmetric Reactions: Historical Aspects and Recent Examples. *Int. J. Quant. Chem.* **2015,** *115*, 1232-1249.
5. Lam, Y.; Grayson, M. N.; Holland, M. C.; Simon, A.; Houk, K. N., Theory and Modeling of Asymmetric Catalytic Reactions. *Acc. Chem. Res.* **2016,** *49*, 750-762.
6. Maji, R.; Mallojjala, S. C.; Wheeler, S. E., Chiral Phosphoric Acid Catalysis: From Numbers to Insights. *Chem. Soc. Rev.* **2018**.
7. Jindal, G.; Sunoj, R. B., Rational Design of Catalysts for Asymmetric Diamination Reaction Using Transition State Modeling *Org. Biomol. Chem.* **2014,** *12*, 2745-2753.
8. Seguin, T. J.; Wheeler, S. E., Stacking and Electrostatic Interactions Drive the Stereoselectivity of Silylium-Ion Asymmetric Counteranion-Directed Catalysis. *Angew. Chem. Int. Ed.* **2016,** *55*, 15889-15893.
9. Straker, R. N.; Peng, Q.; Mekareeya, A.; Paton, R. S.; Anderson, E. A., Computational Ligand Design in Enantio- and Diastereoselective Ynamide [5+2] Cycloisomerization. *Nat. Commun.* **2016,** *7*, 10109.
10. Nimmagadda, S. K.; Mallojjala, S. C.; Woztas, L.; Wheeler, S. E.; Antilla, J. C., Enantioselective Synthesis of Chiral Oxime Ethers: Desymmetrization and Dynamic Kinetic Resolution of Substituted Cyclohexanones. *Angew. Chem. Int. Ed.* **2017,** *56*, 2454-2458.
11. DiRocco, D. A.; Ji, Y.; Sherer, E. C.; Klapars, A.; Reibarkh, M.; Dropinski, J.; Mathew, R.; Maligres, P.; Hyde, A. M.; Limanto, J.; Brunskill, A.; Ruck, R. T.; Campeau, L.-C.; Davies, I. W., A Multifunctional Catalyst That Stereoselectively Assembles Prodrugs. *Science* **2017,** *356*, 426-430.
12. Guan, Y.; Wheeler, S. E., Automated Quantum Mechanical Predictions of Enantioselectivity in a Rhodium-Catalyzed Asymmetric Hydrogenation. *Angew. Chem. Int. Ed.* **2017,** *56*, 9101-9105.
13. Seeman, J. I., Effect of Conformational Change on Reactivity in Organic Chemistry. Evaluations, Applications, and Extensions of Curtin-Hammett Winstein-Holness Kinetics. *Chem. Rev.* **1983,** *83*, 83-134.
14. Lu, T.; Porterfield, M. A.; Wheeler, S. E., Explaining the Disparate Stereoselectivities of N-Oxide Catalyzed Allylations and Propargylations of Aldehydes. *Org. Lett.* **2012,** *14*, 5310-5313.

15. Zou, Y.; Gutierrez, O.; Sader, A. C.; Patel, N. D.; Fandrick, D. R.; Busacca, C. A.; Fandrick, K. R.; Kozlowski, M.; Senanayake, C. H., A Computational Investigation of the Ligand-Controlled Cu-Catalyzed Site-Selective Propargylation and Allenylation of Carbonyl Compounds. *Org. Lett.* **2017**.
16. Schlegel, H. B., Optimization of Equilibrium Geometries and Transition Structures. *J. Comput. Chem.* **1982,** *3*, 214-218.
17. Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J., Using Redundant Internal Coordinates to Optimize Equilibrium Geometries and Transition States. *J. Comput. Chem.* **1996,** *17*, 49-56.
18. Peng, C.; Schlegel, H. B., Combining Synchronous Transit and Quasi‐Newton Methods to Find Transition States. *Israel J. Chem.* **1993,** *33*, 449-454.
19. Irikura, K. K.; Johnson, R. D., Predicting Unexpected Chemical Reactions by Isopotential Searching. *J. Phys. Chem. A* **2000,** *104*, 2191-2194.
20. Zimmerman, P. M., Single-Ended Transition State Finding with the Growing String Method. *J. Comput. Chem.* **2015,** *36*, 601-611.
21. Varela, J. A.; Vázquez, S. A.; Martínez-Núñez, E., An Automated Method to Find Reaction Mechanisms and Solve the Kinetics in Organometallic Catalysis. *Chem. Sci.* **2017,** *8*, 3843-3851.
22. Maeda, S.; Ohno, K., A New Global Reaction Route Map on the Potential Energy Surface of H2co with Unrestricted Level. *Chem. Phys. Lett.* **2008,** *460*, 55-58.
23. Ohno, K.; Maeda, S., Global Reaction Route Mapping on Potential Energy Surfaces of Formaldehyde, Formic Acid, and Their Metal-Substituted Analogues. *J. Phys. Chem. A* **2006,** *110*, 8933-8941.
24. Schlegel, H. B., Following Gradient Extremal Paths. *Theor. Chim. Acta* **1992,** *83*, 15-20.
25. Bondensgård, K.; Jensen, F., Gradient Extremal Bifurcation and Turning Points: An Application to the H2co Potential Energy Surface. *J. Chem. Phys.* **1996,** *104*, 8025-8031.
26. Dallos, M.; Lischka, H.; Ventura Do Monte, E.; Hirsch, M.; Quapp, W., Determination of Energy Minima and Saddle Points Using Multireference Configuration Interaction Methods in Combination with Reduced Gradient Following: The S0 Surface of H2co and the T1 and T2 Surfaces of Acetylene. *J. Comput. Chem.* **2002,** *23*, 576-583.
27. Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A., A Growing String Method for Determining Transition States: Comparison to the Nudged Elastic Band and String Methods. *J. Chem. Phys.* **2004,** *120*, 7877-7886.
28. Zimmerman, P., Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theory Comput.* **2013,** *9*, 3043-3050.
29. Sheppard, D.; Terrell, R.; Henkelman, G., Optimization Methods for Finding Minimum Energy Paths. *J. Chem. Phys.* **2008,** *128*, 134106.
30. Henkelman, G.; Uberuaga, B. P.; Jónsson, H., A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000,** *113*, 9901-9904.
31. Goodrow, A.; Bell, A. T.; Head-Gordon, M., Development and Application of a Hybrid Method Involving Interpolation and Ab Initio Calculations for the Determination of Transition States. *J. Chem. Phys.* **2008,** *129*, 174109.
32. Sameera, W. M. C.; Maeda, S.; Morokuma, K., Computational Catalysis Using the Artificial Force Induced Reaction Method. *Acc. Chem. Res.* **2016,** *49*, 763-773.

33. Maeda, S.; Morokuma, K., Communications: A Systematic Method for Locating Transition Structures of a+B→X Type Reactions. *J. Chem. Phys.* **2010,** *132*, 241102.
34. Maeda, S.; Morokuma, K., Finding Reaction Pathways of Type a + B → X: Toward Systematic Prediction of Reaction Mechanisms. *J. Chem. Theory Comput.* **2011,** *7*, 2335-2345.
35. Maeda, S.; Morokuma, K., Toward Predicting Full Catalytic Cycle Using Automatic Reaction Path Search Method: A Case Study on Hco(Co)3-Catalyzed Hydroformylation. *J. Chem. Theory Comput.* **2012,** *8*, 380-385.
36. Maeda, S.; Ohno, K.; Morokuma, K., Systematic Exploration of the Mechanism of Chemical Reactions: The Global Reaction Route Mapping (Grrm) Strategy Using the Addf and Afir Methods. *Phys. Chem. Chem. Phys.* **2013,** *15*, 3683-3701.
37. Bhoorasingh, P. L.; Slakman, B. L.; Seyedzadeh Khanshan, F.; Cain, J. Y.; West, R. H., Automated Transition State Theory Calculations for High-Throughput Kinetics. *J. Phys. Chem. A* **2017,** *121*, 6896-6904.
38. Bhoorasingh, P. L.; West, R. H., Transition State Geometry Prediction Using Molecular Group Contributions. *Phys. Chem. Chem. Phys.* **2015,** *17*, 32173-32182.
39. Jacobson, L. D.; Bochevarov, A. D.; Watson, M. A.; Hughes, T. F.; Rinaldo, D.; Ehrlich, S.; Steinbrecher, T. B.; Vaitheeswaran, S.; Philipp, D. M.; Halls, M. D., Automated Transition State Search and Its Application to Diverse Types of Organic Reactions. *J. Chem. Theory Comput.* **2017,** *13*, 5780-5797.
40. Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M., Methods for Exploring Reaction Space in Molecular Systems. *WIREs Comput. Mol. Sci.* **2018,** *8*, e1354.
41. Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A., Jaguar: A High‐Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *Int. J. Quantum Chem.* **2013,** *113*, 2110-2142.
42. Duarte, F.; Paton, R. S., Molecular Recognition in Asymmetric Counteranion Catalysis: Understanding Chiral Phosphate-Mediated Desymmetrization. *J. Am. Chem. Soc.* **2017,** *139*, 8886-8896.
43. Medvedev, M. G.; Panova, M. V.; Chilov, G. G.; Bushmarinov, I. S.; Novikov, F. N.; Stroganov, O. V.; Zeifman, A. A.; Svitanko, I. V., Exhaustive Conformational Search for Transition States: The Case of Catechol O-Methyltransferase Active Site. *Mendeleev Commun.* **2017,** *27*, 224-227.
44. Ardkhean, R.; Roth, P. M.; Maksymowicz, R. M.; Curran, A.; Peng, Q.; Paton, R. S.; Fletcher, S. P., Enantioselective Conjugate Addition Catalyzed by a Copper Phosphoramidite Complex: Computational and Experimental Exploration of Asymmetric Induction. *ACS Catal.* **2017,** *7*, 6729-6737.
45. Gammack Yamagata, A. D.; Datta, S.; Jackson, K. E.; Stegbauer, L.; Paton, R. S.; Dixon, D. J., Enantioselective Desymmetrization of Prochiral Cyclohexanones by Organocatalytic Intramolecular Michael Additions to A,B-Unsaturated Esters. *Angew. Chem. Int. Ed.* **2015,** *54*, 4899-4903.
46. Chang, G.; Guida, W. C.; Still, W. C., An Internal Coordinate Monte Carlo Method for Searching Conformational Space. *J. Am. Chem. Soc.* **1989,** *1111*, 4379-4386.
47. Minenkov, Y.; Sharapa, D. I.; Cavallo, L., Application of Semiempirical Methods to Transition Metal Complexes: Fast Results but Hard-to-Predict Accuracy. *J. Chem. Theory Comput.* **2018**.

48. AARON: An Automated Reaction Optimizer for New catalysts, v. 1.0, Y. Guan, V. M. Ingman, B. J. Rooks, and S. E. Wheeler, Texas A&M University, College Station, TX, 2018. https://github.com/QChASM/Aaron
49. Available from https://github.com/QChASM/AaronTools
50. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminformatics* **2012,** *4*, 17.
51. Ioannidis, E. I.; Gani, T. Z.; Kulik, H. J., Molsimplify: A Toolkit for Automating Discovery in Inorganic Chemistry. *J. Comput. Chem.* **2016,** *37*, 2106-2117.
52. RDKit: Open-Source Cheminformatics Software. http://www.rdkit.org/
53. Currently, AaronTools is primarily geared toward Gaussian 09 and Gaussian 16, but is being extended to Orca 4.0.
54. Guan, Y.; Wheeler, S. E., Intercolumnar Interactions Control the Local Orientations within Columnar Stacks of Sumanene and Sumanene Derivatives. *The Journal of Physical Chemistry C* **2017,** *121*, 8541-8547.
55. 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.
56. Currently available ligands, organocatalysts, and substituents can be viewed at http://catalysttrends.org/AaronTools
57. Wheeler, S. E.; Houk, K. N., Integration Grid Errors for Meta-Gga-Predicted Reaction Energies: Origin of Grid Errors for the M06 Suite of Functionals. *Journal of Chemical Theory and Computation* **2010,** *6*, 395-404.
58. Currently available TS templates can be viewed at http://catalysttrends.org/Aaron
59. Lu, G.; Fang, C.; Xu, T.; Dong, G.; Liu, P., Computational Study of Rh-Catalyzed Carboacylation of Olefins: Ligand-Promoted Rhodacycle Isomerization Enables Regioselective C–C Bond Functionalization of Benzocyclobutenones. *J. Am. Chem. Soc.* **2015,** *137*, 8274-8283.
60. Kearsley, S. K., On the Orthogonal Transformation Used for Structural Comparisons. *Acta Cryst.* **1989,** *A45*, 208-210.
61. Wang, S.-G.; Liu, X.-J.; Zhao, Q.-C.; Zheng, C.; Wang, S.-B.; You, S.-L., Asymmetric Dearomatization of B-Naphthols through a Bifunctional-Thiourea-Catalyzed Michael Reaction. *Angew. Chem. Int. Ed.* **2015,** *54*, 14929-14932.
62. Some conformational changes must still be manually added to the TS template library, including macrocycle conformations and ring-flips.
63. Hawkins, P. C. D., Conformation Generation: The State of the Art. *J. Chem. Inf. Model.* **2017,** *57*, 1747-1756.
64. Grimme, S., Supramolecular Binding Thermodynamics by Dispersion‐Corrected Density Functional Theory. *Chem. Eur. J.* **2012,** *18*, 9955-9964.
65. Werner, E. W.; Mei, T.-S.; Burckle, A. J.; Sigman, M. S., Enantioselective Heck Arylations of Acyclic Alkenyl Alcohols Using a Redox-Relay Strategy. *Science* **2012,** *338*, 1455-1458.
66. Zhang, C.; Santiago, C. B.; Crawford, J. M.; Sigman, M. S., Enantioselective Dehydrogenative Heck Arylations of Trisubstituted Alkenes with Indoles to Construct Quaternary Stereocenters. *J. Am. Chem. Soc.* **2015,** *137*, 15668-15671.
67. Patel, H. H.; Sigman, M. S., Palladium-Catalyzed Enantioselective Heck Alkenylation of Acyclic Alkenols Using a Redox-Relay Strategy. *J. Am. Chem. Soc.* **2015,** *137*, 3462-3465.

68. Mei, T.-S.; Patel, H. H.; Sigman, M. S., Enantioselective Construction of Remote Quaternary Stereocentres. *Nature* **2014,** *508*, 340.
69. Xu, L.; Hilton, M. J.; Zhang, X.; Norrby, P.-O.; Wu, Y.-D.; Sigman, M. S.; Wiest, O., Mechanism, Reactivity, and Selectivity in Palladium-Catalyzed Redox-Relay Heck Arylations of Alkenyl Alcohols. *J. Am. Chem. Soc.* **2014,** *136*, 1960-1967.
70. Zhang, C.; Tutkowski, B.; DeLuca, R. J.; Joyce, L. A.; Wiest, O.; Sigman, M. S., Palladium-Catalyzed Enantioselective Heck Alkenylation of Trisubstituted Allylic Alkenols: A Redox-Relay Strategy to Construct Vicinal Stereocenters. *Chem. Sci.* **2017,** *8*, 2277-2282.
71. Knowles, W. S., Asymmetric Hydrogenations (Nobel Lecture). *Angew. Chem. Int. Ed.* **2002,** *41*, 1998-2007.
72. Landis, C. R.; Hilfenhaus, P.; Feldgus, S., Structures and Reaction Pathways in Rhodium(I)-Catalyzed Hydrogenation of Enamides: A Model Dft Study. *J. Am. Chem. Soc.* **1999,** *121*, 8741-8754.
73. Landis, C. R., A Simple Model for the Origin of Enantioselection and the Anti "Lock‐and‐Key" Motif in Asymmetric Hydrogenation of Enamides as Catalyzed by Chiral Diphosphine Complexes of Rh (I). *Angew. Chem. Int. Ed.* **2000,** *39*, 2863-2866.
74. Feldgus, S.; Landis, C. R., Origin of Enantioreversal in the Rhodium-Catalyzed Asymmetric Hydrogenation of Prochiral Enamides and the Effect of the A-Substituent. *Organometallics* **2001,** *20*, 2374-2386.
75. Donoghue, P. J.; Helquist, P.; Wiest, O., Ligand and Substrate Effects on the Mechanism of Rhodium-Catalyzed Hydrogenation of Enamides. *The Journal of organic chemistry* **2007,** *72*, 839-847.
76. Liu, G.; Liu, X.; Cai, Z.; Jiao, G.; Xu, G.; Tang, W., Design of Phosphorus Ligands with Deep Chiral Pockets: Practical Synthesis of Chiral B‐Arylamines by Asymmetric Hydrogenation. *Angew. Chem. Int. Ed.* **2013,** *52*, 4235-4238.
77. 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 Catal.* **2014,** *5*, 272-280.
78. Doney, A. C.; Rooks, B. J.; Lu, T.; Wheeler, S. E., Design of Organocatalysts for Asymmetric Propargylations through Computational Screening. *ACS Catal.* **2016,** *6*, 7948-7955.
79. Sepúlveda, D.; Lu, T.; Wheeler, S. E., Performance of Dft Methods and Origin of Stereoselectivity in Bipyridine N, N′-Dioxide Catalyzed Allylation and Propargylation Reactions. *Org. Biomol. Chem.* **2014,** *12*, 8346-8353.
80. Lu, T.; Zhu, R.; An, Y.; Wheeler, S. E., Origin of Enantioselectivity in the Propargylation of Aromatic Aldehydes Catalyzed by Helical N-Oxides. *J. Am. Chem. Soc.* **2012,** *134*, 3095-3102.
81. Previous experience applying AARON indicates that the structures in the TS template library do not necessarily need to be fully optimized transition states—AARON is adept at locating saddle-points even starting from relatively poor initial guesses. Practically, one can start with a relatively crude TS template library (either generated at a very low level of theory or constructed with a graphical molecular builder) and then use AARON to construct a more precise library that is then used for production-level predictions.
82. Zimmerman, P. M., Automated Discovery of Chemically Reasonable Elementary Reaction Steps. *J. Comput. Chem.* **2013,** *34*, 1385-1392.

83. Pendleton, I. M.; Pérez-Temprano, M. H.; Sanford, M. S.; Zimmerman, P. M., Experimental and Computational Assessment of Reactivity and Mechanism in C(Sp3)–N Bond-Forming Reductive Elimination from Palladium(Iv). *J. Am. Chem. Soc.* **2016,** *138*, 6049-6060.
84. Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A., Complex Chemical Reaction Networks from Heuristics-Aided Quantum Chemistry. *J. Chem. Theory Comput.* **2014,** *10*, 897-907.
85. Bergeler, M.; Simm, G. N.; Proppe, J.; Reiher, M., Heuristics-Guided Exploration of Reaction Mechanisms. *J. Chem. Theory Comput.* **2015,** *11*, 5712-5722.
86. Suleimanov, Y. V.; Green, W. H., Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods. *J. Chem. Theory Comput.* **2015,** *11*, 4248-4259.
87. Lee, J.; Lee, I.-H.; Joung, I.; Lee, J.; Brooks, B. R., Finding Multiple Reaction Pathways Via Global Optimization of Action. *Nature Communications* **2017,** *8*, 15443.
88. Santiago, C. B.; Guo, J.-Y.; Sigman, M. S., Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018,** *9*, 2398-2412.
89. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.; Izmaylov, A.; Bloino, J.; Zheng, G.; Sonnenberg, J.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., JA; Peralta, J.; Ogliaro, F.; Bearpark, M.; Heyd, J.; Brothers, E.; Kudin, K.; Staroverov, V.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J.; Klene, M.; Knox, J.; Cross, J.; Bakken, V.; Adamo, C.; ramillo, J.; Gomperts, R.; Stratmann, R.; Yazyev, O.; Austin, A.; Cammi, R.; Pomelli, C.; Ochterski, J.; Martin, R.; Morokuma, K.; Zakrzewski, V.; Voth, G.; Salvador, P.; Dannenberg, J.; Dapprich, S.; Daniels, A.; Farkas, O.; Foresman, J.; Ortiz, J.; Cioslowski, J.; Fox, D. *Gaussian 09, Revision D.01*, Gaussian, Inc.: 2009.

TOC Figure