AUTOMATING QUANTUM MECHANICAL PREDICTIONS OF THE STEREOSELECTIVITY OF ASYMMETRIC REACTIONS AND PROGRESS TOWARD UNDERSTANDING THE PACKING OF CURVED POLYCYCLIC AROMATIC HYDROCARBONS

A Dissertation

by

YANFEI GUAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Steven E. Wheeler
Co-Chair of Committee,	Michael B. Hall
Committee Members,	Lei Fang
	Rainer J. Fries
Head of Department,	Simon W. North

August 2018

Major Subject: Chemistry

ABSTRACT

Modern computational chemistry has emerged as a powerful tool for understanding functional molecules including catalysts, materials, and pharmaceuticals. Such theoretical studies can inform the rational design of new structures such as effective catalysts and high performed organic semiconductors. In this dissertation, applications of computational chemistry to the fields of asymmetric catalysis and organic electronic materials based on polycyclic aromatic hydrocarbons are presented.

The first part of this dissertation focuses on the development and application of computational tools for the automated prediction of stereoselectivities for asymmetric catalytic reactions. Despite the enormous success of quantum chemistry in explaining the activity and selectivity of a wide range of catalytic reactions, the computational design of new catalysts is far from routine due to the vast number of computations that must be performed for a given system. A computational toolkit has been developed (AARON: An Automated Reaction Optimizer for New-catalysts) to automatically screen virtual libraries of potential catalyst designs quantum mechanically. First, an explanation of the AARON workflow is provided along with several representative applications demonstrating the utility of AARON in the context of catalysts design. Then, a more detailed example of the application of AARON to a Rh-catalyzed asymmetric hydrogenation reaction is presented in which AARON is used to screen a series of ligands and design a novel chiral.

In the second part, the primary goal is to understand the underlying noncovalent interactions responsible for the supramolecular assembly of curved molecules and thus to precisely control supramolecular architectures. First, a study of four bowl-shaped polycyclic aromatic hydrocarbon molecules (summanene and derivatives) is presented that combines quantum mechanical methods and molecular dynamics simulations to understand the local orientations within columnar stacks, which significantly impacts the electronic properties of the resulting semiconducting materials. The importance of intercolumnar interactions on intracolumnar packing is also highlighted. Next, we consider contorted polycyclic aromatic molecules. These doubly-concave systems offer a means of tuning the strength of stacking interactions through variations in molecular curvature. A systematically study of a set of 18 saddle-shaped compounds and their dimers indicates that the introduction of thiophene rings around the periphery of these systems and the incorporation of B and N atoms into the coronene core greatly enhances their tendency to form strongly stacked dimers. Ultimately, a complex relationship between curvature of molecules and stacking interaction is revealed.

Together, the chapters in this thesis showcase the power of modern computational quantum chemistry to provide useful predictions of the structure and properties of molecules.

DEDICATION

To my wonderful girlfriend, Yaqi Dai, for all her love and support.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor and committee chair, Dr. Wheeler, for all his help and supervision on my study and research. Dr. Wheeler is an excellent professor and mentor, who led me to the world of computational chemistry and taught me how to be a qualified researcher. I really enjoyed working with him. I would like to thank my committee co-chair, Dr. Hall, for providing me invaluable advices on computational methods. I am also grateful to Dr. Fang and Dr. Fries for serving on my committee.

I would like to thank all of my professors in the Department of Chemistry at Texas A&M through the past four years who have been involved in my chemical education, Dr. North in particular. Also thanks to Sandy Manning and Julie Zercher for helping me meet the requirements for the Ph.D. program on time.

Thanks to my great and helpful co-workers in Wheeler group for their valuable discussions which often break bottlenecks in my research. Thanks to every one of Center for Computational Quantum Chemistry, University of Georgia, Dr. Schaeffer and Dr. Allen in particular, for the more theoretical knowledge I learned there. Thanks to the Jaguar development team at Schrödinger, Inc, where I had an unforgettable summer and improved my programming skill a lot. The experience at Schrödinger triggered many interesting ideas in the development of AARON.

Finally, I would like to thank my girlfriend. We met at Texas A&M. It was her love and encouragement that support me through my Ph.D. study. I am also grateful to my parents and aunt for their encouragement and support.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Steven E. Wheeler (advisor and committee chair), Professor Michael B. Hall (co-committee chair), and Professor Lei Fang of the Department of Chemistry, and Professor Rainer J. Fires of the Department of Physics & Astronomy.

Experimental data in Chapters III were published previously by Professor Wenjun Tang.

All other work in the dissertation was conducted independently by the student under the supervision of Steven E. Wheeler.

Funding Sources

The work in this dissertation was supported by the National Science Foundation (Grants CHE-1266022, CHE-1665407, and CHE-1254897).

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Computational chemistry is a branch of chemistry that uses theoretical methods to compute the properties of molecules or solids. These methods are usually incorporated into efficient computational packages and can be based on either classical- or quantum-mechanical treatments of molecular systems. Examples of molecular properties that can be computed include structures, relative energies, vibrational frequencies, dipoles and higher multipole moments, and other spectroscopic properties. The prediction of these properties using computational chemistry has found increasing applications in areas ranging from materials chemistry and catalyst design to drug discovery, *etc*.

Modern computational chemistry methods fall into two large categories, classical-mechanical and quantum-mechanical methods, each containing many different approaches designed for different purposes. These computational methods range from very approximate but highly computationally-efficient coarse-grained models to highly accurate but computationally expensive correlated *ab initio* methods. In all cases, the computer time and other resources required (such as CPU/GPU time and memory) increase rapidly with the size of system being studied. In general, higher accuracy can be achieved by using higher levels of theory at the cost of more computational time. As a result, methods delivering high accuracy predictions are often feasible only for small systems.

With our concentration on small molecule catalysts and non-covalent interactions in organic materials, we focus on density functional theory (DFT) and molecular mechanical (MM) methods. These families of methods are described briefly below.

1.1 DFT Methods for Non-Covalent Interactions and Homogeneous Catalysis

DFT is an important family of quantum-mechanical methods for investigating the electronic structure of many-body systems. Since its introduction into computational chemistry in the 1970s, DFT has emerged as one of the most popular and versatile methods available for practical applications of computational chemistry to problems in homogeneous catalysis and intermolecular interactions due to its excellent balance of accuracy and computational cost.

Quantum chemical methods are primarily aimed at solving the time-independent Schrödinger equation,

$$[T + V(\boldsymbol{r})]\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

where *T* and *V(r)* are the kinetic and potential energy operators, respectively, $\Psi(r)$ is the (time-independent) stationary state wavefunction, and E is the total energy of the system. Applied to molecules under the Born-Oppenheimer (BO) approximation, the nuclear kinetic energy is neglected and electrons move under the Coulombic potential of the nuclei at fixed positions in space. This results in the electronic Schrödinger equation,

$$[T_e + V_e(\boldsymbol{r}, \boldsymbol{R})]\chi(\mathbf{r}, \mathbf{R}) = E_e\chi(\mathbf{r}, \mathbf{R})$$

where T_e and $V_e(r, R)$ are the kinetic and potential energy operators for electrons, respectively, and $\chi(r, R)$ is the electronic wavefunction when the nuclei are at positions **R**.

Varying the positions of the nuclei and repeatedly solving the electronic Schrödinger equation yields E_e as a function of the nuclear coordinates (**R**), which is the potential energy surface (PES) on which the nuclei move.

Since the electronic Schrödinger equation cannot be solved analytically for systems with multiple electrons due to electron-electron interactions, approximations must be made to obtain estimated wavefunctions and energies for a given molecular system. Many efforts have been made to provide approximate solutions to the electronic Schrödinger equation. The most fundamental approach dates back to the pioneering work of Hartree and Fock in the 1920s (the HF method). In HF theory, the electronic wavefunction is approximated by a single Slater determinant of N spin-orbitals. By the variational theorem, the HF energy is an upper-bound to the true ground state electronic energy. By iteratively solving the HF equations and thus optimizing the HF orbitals until a self-consistent field (SCF) is achieved, one can derive an approximate wavefunction and energy for a molecular system. One major shortcoming of HF is that it is a mean field approximation, in which each electron is subjected to the mean Coulombic field of the N-1 other electrons. That is, in HF theory electrons move independently of each other, neglecting 'electron correlation'. This neglect of electron correlation can lead to large deviations between predictions from HF theory and experimental results.

Starting from the HF method, a number of more elaborate post-Hartree-Fock quantum chemical approaches have been devised to recover electron correlation, such as Møller–Plesset (MP) perturbation theory, configuration interaction (CI), and coupled-cluster theory. Coupled-cluster theory with single, double, and perturbative triple

excitations, CCSD(T), has been established as the "gold standard" for first principles computations in quantum chemistry and applied widely to study small molecules. Unfortunately, these post-HF approaches usually present difficulties in applications to systems containing more than ~20 atoms due to the sharp increase in computational cost. For example, the CCSD(T) method scales as $O(N^7)$, where N is representative of the system size.

By replacing the many-body electronic wavefunction with the electronic density, DFT attempts to address both the inaccuracy of HF methods and high computational cost of post-HF methods. Modern DFT rests on two theorems by Pierre Hohenberg and Walter Kohn (H-K). The first H-K theorem states that the ground state properties of a many-electron system are uniquely determined by an electron density. Whereas the many-electron wavefunctions appearing in post-HF methods depend on $3N$ variables for an N electron system, the electron density depends only on three variables. This means that DFT is simpler both conceptually and practically, with formal scaling of $O(N^3)$ to $O(N^4)$. The second H-K theorem states that the energy of an electronic system can be described by a functional of the electron density function, and the correct electron density minimizes the value of this energy functional.

Although those two H-K theorems assure us the existence of this energy functional, unfortunately the exact form of such a universal functional is unknown. From the two H-K theorems, we can write the total electronic energy as

$$E[\rho] = E_{Ne}[\rho] + T[\rho] + E_{ee}[\rho]$$

where only the form of nuclear attraction term, $E_{Ne}[\rho]$, is known. If we define

$$F_{HK}[\rho] = T[\rho] + E_{ee}[\rho]$$

We have

$$E[\rho] = \int \rho(r) V_{Ne(r)} dr + F_{HK}[\rho]$$

This functional $F_{HK}[\rho]$ is the holy grail of DFT and if it were known we would have a computationally inexpensive route to exact solutions of the electronic Schrödinger equation for all molecular systems. $F_{HK}[\rho]$ contains contributions from the kinetic energy and electron-electron interaction potential energy, which can be further written as sum of classical Coulomb interactions and a non-classical classical component:

$$F_{HK}[\rho] = T[\rho] + J[\rho] + E_{ncl}[\rho]$$

In practice, current common DFT methods are based on the formulation by Walter Kohn and Lu Jeu Sham in what is known as Kohn-Sham (K-S) DFT. Borrowing the concept of orbitals from HF methods, in K-S DFT a fictitious noninteracting system is constructed to yield the same electron density as the true system of interacting electrons. Since this non-interacting system only consists of noninteracting electrons, the wavefunction can be exactly represented as a Slater determinant of orbitals, and the kinetic energy of the noninteracting system can be computed exactly as

$$T_S = -\frac{1}{2}\sum \langle \varphi | \nabla^2 | \varphi \rangle$$

However, T_S is not equal to the true kinetic energy of the interacting system. As a result, a non-classical component is introduced to account for that,

$$E_{xc}[\rho] = (T[\rho] - T_S[\rho]) + E_{ncl}[\rho]$$

where the exchange-correlation energy, $E_{xc}[\rho]$, contains everything unknown in K-S DFT. Given this, we have the total energy for the K-S DFT expression

$$E[\rho] = E_{Ne}[\rho] + T[\rho] + J[\rho] + E_{xc}[\rho]$$

The exchange-correlation energy is where DFT differs from the HF method. The correlation energy arises from the correlated motion of interacting electrons and is defined as the difference between exact energy and the HF energy. The exchange component reflects the exchange interaction that arises from the requirement that wavefunctions of indistinguishable particles must exhibit exchange symmetry. Together, the exchange functional and correlation functional define a given DFT functional. As with HF theory, the Kohn-Sham equations must be solved self-consistently to obtain optimized K-S orbitals and a minimized K-S energy.

The simplest form for an exchange-correlation functional is the local-density approximation (LDA), in which the energy is only determined by the density at each point in space. As the simplest DFT functional, LDA can only deliver moderate accuracy, which is insufficient for most chemical applications. The breakthrough of DFT came with the introduction of generalized gradient approximation (GGA) functionals, which include information about the gradient of electron density to better mirror the inhomogeneous nature of the electron density. Using GGAs, such as the BP86 and PBE functionals, accurate molecular geometries and ground-state energies can be obtained. As an improvement to GGAs, hybrid GGA functionals were introduced that incorporate a portion of 'exact exchange' energy from the HF method. For example, in the B3LYP functional 20% of exchange energy comes from a HF-like expression. Nowadays, hybrid

GGA functionals such as B3LYP are among the most popular functionals for studying organic complex systems with 50-100s atoms. Potentially more accurate functionals are the meta-GGA and hybrid meta-GGA functionals, which include second or higher derivatives of electron density. This includes functionals such as TPSS and the M06 family of functionals, which can yield very accurate energetic and spectroscopic properties.

Although many popular DFT methods proved successful in treating covalently-bound molecules as well as hydrogen bonding and other electrostatically dominated non-covalent interactions, many of these functionals have proved incapable of describing the R^{-6} distance-dependence of dispersion interactions. Dispersion interactions drive many long-range non-covalent interactions that are central to the molecular systems in the present work. The simplest approach to capturing dispersion interaction is the semi-empirical dispersion correction method of Grimme and co-workers (DFT-D), which appends a damped empirical atom-atom pairwise dispersion term to the total DFT energy. In the second-generation DFT-D method (DFT-D2, often written simply as DFT-D), the dispersion correction is C_6 based and proportional to R^{-6}, in which R is the distance between pairs of atoms. In 2006, Antony and Grimme[1] evaluated the performance of the D2 dispersion correction appended to various GGA and hybrid DFT functionals across a database of 22 van der Waals (vdW) complexes (the S22 test set). BLYP-D, PBE-D, B97-D, and B3LYP-D all provided interaction energies with average errors less than 1 kcal mol^{-1} or 10% compared to accurate reference values; only six of the 161 DFT functionals tested showed errors exceeding 2 kcal mol^{-1}. The following

year, Jurecka *et al*.[2] tested the damped dispersion term with six DFT functionals against a larger set of 58 vdW complexes, S58. Results were in remarkable agreement with high-level data based on CCSD(T) and even surpass the MP2/cc-PVTZ level of theory.

Recently, Grimme and co-workers[3] introduced a new dispersion correction (DFT-D3) that incorporates an additional R^{-8} term and adjust the functional form of the R^{-6} term and the damping function. Moreover, in DFT-D3 there are additional atom types corresponding to different hybridization states of different elements. Subsequent benchmarks have shown that DFT-D3 yields superior results for many functionals compared to DFT-D2.[4-8] Nowadays, the DFT-D3 method is broadly applied to computational studies of non-covalent interactions, given its extremely good performance/cost ratio.

1.2 Molecular Mechanics Methods for Simulations of Bulk Materials

Although QM methods can be widely applied to medium sized molecular systems containing ~100 atoms, such rigorous approaches are not applicable to systems such as biomolecules and materials. To study these systems, classical-mechanical methods based on empirical molecular mechanics (MM) force fields are necessary.

MM methods, in contrast to QM-based methods that aim to solve the electronic Schrödinger equation, rely on classical type models to predict potential energy of molecular systems as a function of nuclear coordinates. In all-atom MM simulations each atom is treated as a particle, while the impact of the electrons is reflected in atomic properties such as partial charges and polarizabilities. Covalently bonded interactions

are treated as springs with force constants derived from either experimental data or *ab initio* computed force constants and equilibrium geometries. In general, deviation of covalent bonds from their equilibrium geometries and noncovalent interactions give rise to the total interaction potential energy. In MM methods, the covalent terms are comprised of bond, angle, and dihedral contributions while noncovalent interactions involve either electrostatic interactions between atom-based partial charges and Lennard-Jones (LJ) interactions. A typical functional form of the very popular OPLS force field is given by

$$E = \sum_{bonds} k_r(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + E_{dihedrals}(\varphi) + \sum_{i>j} f_{ij} \frac{q_i q_i e^2}{4\pi\varepsilon_0 r_{ij}} + \sum_{i>j} f_{ij} \left(\frac{A_{ij}}{r^{12}} - \frac{C_{ij}}{r^6}\right)$$

where the last two terms are the electrostatic and LJ energies, respectively. Generally, the bond and angle terms are treated as harmonic oscillators; the dihedral terms typically cannot be modeled as harmonic potentials, given that they can have multiple energy minima.

In addition to the OPLS force field, there are other popular MM force fields that differ in either the functional form or the associated force constants. These functional forms and associated force constants as well as equilibrium bond, angle, dihedral values, are collectively termed a force field. Parameterization is typically done through training by a set of experimental or ab initio data.

The main use of molecular mechanics lies in the field of molecular dynamics (MD) simulations, which is also how we use the MM method in the present dissertation. Since the MM force field gives the energy for the system as a function of atomic positions, which allows us to compute the force exerted on each atom given the coordinates of all atoms, we can predict motions of each atom by Newton's second law. In MD simulations, atoms and molecules are allowed to interact for a fixed period of time under a specific force field, which is divided into discrete time steps (typically 2 *fs*). At each time step, the force acting on each atom will be computed to update the position and velocity for the next time step. Doing this iteratively, we obtain a trajectory of the molecular motion. MD simulations can be performed under several statistical mechanical conditions, such as microcanonical ensemble (NVE), canonical ensemble (NVT), isothermal-isobaric (NPT) ensemble, *etc*. In this dissertation, most simulations were conducted under an NPT ensemble, in which the amount of substances (N), pressure (P), and temperature (T) are conserved. It corresponds most closely to the laboratory conditions and thus plays an important role in many chemistry and materials simulations.

If the MD simulation time is long enough, the system will reach an equilibrium state, which typically mirrors the preferred configuration and conformation of the system under the provided pressure and temperature. In the present dissertation, we use MD simulations to determine the external environment of curved polycyclic molecules in bulk materials to better understand the intermolecular interaction in the organic semi-conductive materials.

1.3 Computational design of asymmetric catalysts

Asymmetric catalysis describes the use of a chiral catalyst to accelerate a chemical reaction such that there is preferential formation of one stereoisomer of a chiral product. Driven in large part by the need for chiral compounds of pharmaceutical interest, as well as other applications including fragrances, materials, *etc*., there has been a tremendous surge of interest in asymmetric catalysis.[9] Before the advent of asymmetric catalysis, enantiomerically enriched compounds were generated either by resolving them from a racemic mixture or by the transformation of nature chiral molecules. However, those two approaches are limited by low yields and the limited availability of chiral nature products. A major breakthrough in asymmetric catalysis came in early 1970s, when scientists at Monsanto developed Rh complexes containing chiral phosphine ligands as asymmetric catalyst for the enantioselective addition of H_2 to olefinic substrates.[10] Since then, asymmetric catalysis has dramatically changed the process of generating chiral compounds.[11] When first developed around 1970s, homogenous asymmetric catalysts were based on transition metal complexes consisting of transition metal centers surrounded by chiral ligands.[12] This includes the well-known complex of Rh with DIPAMP ligands, which provides selectivies exceeding 95% ee for the asymmetric hydrogenation reaction and was the first industrial-scale asymmetric catalyst,[13] esp. the synthesis of the drug L-DOPA.[13] Since the 1970s, various chiral ligands have been developed in order to obtain both higher stereoselectivities and wider substrates scopes. Nowadays, excellent selectivities have been achieved with a wide array of transition metal complexes, including Rh, Ru, Cu, Ni, Pd, Ir, *etc*.[14-19]

Despite these achievements, transition metal catalysts are still limited by high price, low availability, toxicity, environmental unfriendliness, and difficulty of use. Since the late 1990s, there has been a trend toward using small metal free organic molecules to carry out enantioselective transformations.[20] Typically, there are four classes of organocatalysts, Lewis bases, Lewis acids, Brønsted bases and Brønsted acids, among which Lewis base catalysts, such as amines and carbenes,[21-24] dominates this area.[25] Interest in chiral Brønsted acid catalysts such as chiral phosphoric acids,[26, 27] for example chiral BINOL phosphate complex,[28] has grown steadily in recent years. These catalysts activate substrates through hydrogen bonding and/or proton transfer and can exhibit extremely high degrees of selectivity that are competitive with transition metal catalysts.[29]

Although the field of homogenous asymmetric catalysis continues to flourish, the process of catalyst design can hardly meet the expanding demand for more effective and selective catalysts since most organic and organometallic catalysts are developed largely through experimental trial and error. Modern computational quantum chemistry has emerged as a potentially powerful tool for rational catalyst design, because it can elucidate the mechanisms of catalytic reactions by locating transition states (TS) and intermediates that are typically hard to detect experimentally.[27, 30-33] Upon computing the free energies of key TSs for all stereoisomeric products using, for example, DFT methods, one can predict selectivities based on the Curtin-Hammett principle,[34] which states that the product distribution reflects the difference in free energy between the two rate-limiting TSs ($\Delta\Delta G^{\ddagger}$ in Figure 1). Consequently, central to the computational study

of asymmetric catalysis is the optimization of structures of TS geometries leading to formation of both the major and minor stereoisomeric products. The selectivity of a given catalytic reaction can be characterized by either enantiomeric ratio (*er*) or enantiomeric excess (*ee*),

$$er = \frac{[R]}{[S]} = e^{\frac{\Delta\Delta G^{\ddagger}}{RT}}$$

$$ee = \frac{[R] - [S]}{[R] + [S]} = \frac{1 - e^{\frac{\Delta\Delta G^{\ddagger}}{RT}}}{1 + e^{\frac{\Delta\Delta G^{\ddagger}}{RT}}}$$

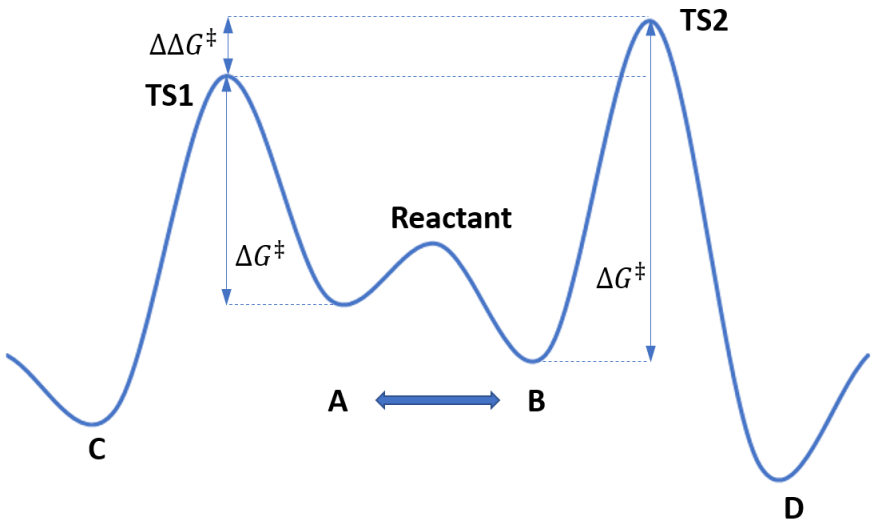


Figure I-1. Curtin-Hammett scenario. Since pre-reaction complex A and B are in rapid equilibrium, the distribution of products C and D only depends on the difference in free energy of the transition states, TS1 and TS2, leading to each product, $\Delta\Delta G^{\ddagger}$.

The success of locating a transition state structure (a first-order saddle-point on the 3*N*-6 dimensional PES, where *N* is the number of atoms) depends on the quality of the initial guess structure provided by computational chemists, and thus depend strongly

on chemical intuition and experience.[35] Because of the complexity of reaction potential energy surfaces, the number of accessible transition states can be enormous.[36, 37] Unfortunately, using typical computational methods one can only obtain one TS structure at a time. In order to locate all low-lying TS structures, computational chemists must repeat the tedious work of building initial structures, preparing input files, submitting computations, checking output files, and comparing final optimized TS structures until all low-lying TS structures have been identified (Figure 2).

Another challenge for computational catalysis stems from the large conformational space spanned by complex ligands and substrates. Typically, the larger and more flexible the ligands and substrates are, the more expansive the conformational space. Often, only a single conformation identified by limited manual conformational searching or guessing is considered when predicting selectivity.[38] This can lead to predicted selectivities that are qualitatively incorrect, hindering both our understanding of existing catalytic reactions and precluding the effective use of computational chemistry in prospective catalyst design.

The tedious process of locating TS structures combined with the need for exhaustive conformational searching makes it difficult to screen virtual libraries of potential chiral ligands computationally while still maintaining sufficient accuracy. Instead, experimental high-throughput screening is the current state-of-the-art route to new catalysts.[39-41]

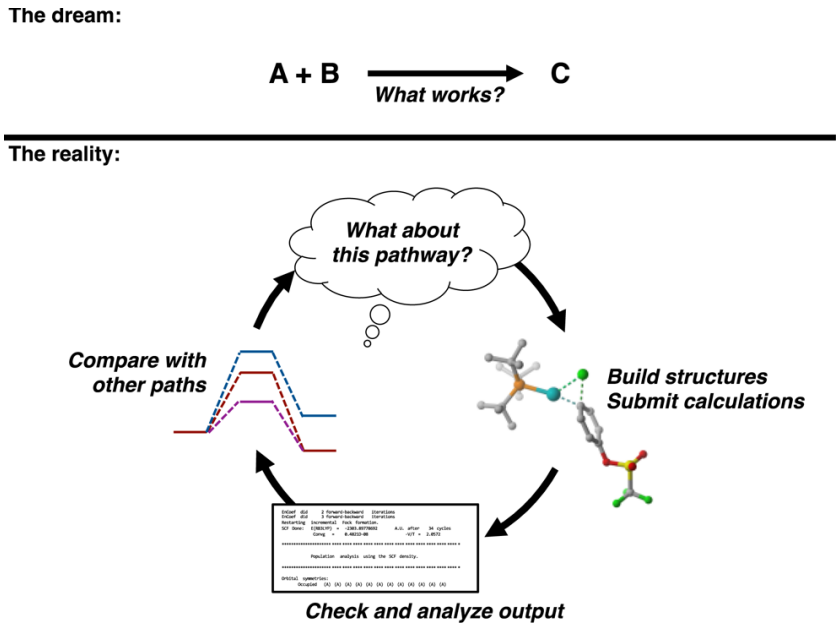


Figure I-2. Conceptual representation of reactivity evaluation using computational methods. Modified with permission from *Accounts of Chemical Research* 2017, *50*, 605-608. Copyright 2017 American Chemical Society.

There have long been attempts to develop methods to automatically locate TS structures to enable computer-aided catalysts design. One such approach is the use the MM-based methods instead of quantum mechanical methods. However, as discussed above, in traditional force fields bonding interactions are treated as harmonic oscillators and hence are not capable of describing saddle points on potential energy surfaces. To describe transition states, transition state force fields (TSFFs) were introduced, which represents TS structures as energy minima. Such TSFFs allow fast conformational sampling of the region of the PES surrounding the TS.[42] However, the application of TSFFs requires a specific set of force-field parameters for each new reaction. To rapidly generate accurate FF parameters, Wiest *et al*.[38, 43] developed quantum-guided molecular

mechanics (Q2MM), which has been applied successfully to the virtual ligand screening of a range of transition metal catalysis.

Other groups have developed automated QM-based methods for locating TS structures. [44-46] In Chapters II and III, we report our efforts to automate the QM-based geometry optimizations of TS structures needed to predict catalytic activities and selectivities. To that end, we have developed an open source computational toolkit (AARON: An Automated Reaction Optimizer for New catalysts) and an object-oriented toolkit to manipulate complex molecular structures.[47] In Chapter II, we introduce AARON, including explanation of the overall workflow and several applications of AARON. In Chapter III, we provide a more detailed example application of AARON to a Rh catalyzed asymmetric hydrogenation reaction.

1.4 Non-covalent Interactions in Curved Polycyclic Aromatic Hydrocarbons and Derivatives

Polycyclic aromatic hydrocarbons (PAHs) are abundant across the universe, being found in abundance in coal and tar deposits, in residues of incomplete combustion of organic materials, and in interstellar clouds. Typically, PAHs can be regarded as two-dimensional open-end segments of pure carbon molecules. For example, the hexa-peri-hexabenzocoronene (HBC), the so-called "super-benzene", can be viewed as segment of graphene, while sumanene and corannulene are treated as pieces of fullerenes. As segments of graphene or fullerenes, PAHs also show remarkable properties with potential applications in the field of organic electronics.[48, 49] In particular, PAHs can be

fabricated into electrostatic and optoelectronic devices[50] such as solar cells,[51, 52] field-effect transistors (FETs),[53-55] and light-emitting diodes.[56] Many factors can affect the performance of these devices, including the formation, transportation, and recombination of charges, *etc*. One of the most important factors in device performance is the mobility of charge carriers.

Charge carrier mobility defines the moving speed of charge carrier in the material in a given direction. To be competitive with silicon materials, charge carrier mobilities for PAH-based materials must approach 1 cm^2/Vs. Besides the intrinsic electronic properties of the materials, these charge carrier mobilities also depends strongly on the macroscopic order of the PAH molecule in the solid state.

Many PAHs, especially small PAHs such as HBC and sumanene, organize in columnar structures through self-assembly in bulk materials, where molecules stack on top of another molecule into columns. Those columns can then arrange in a regular lattice,[57] which have the ability to conduct charges along the stacks of aromatic cores thorough thermally activated hopping of charges. The higher the hopping rate is, the higher the charge carrier mobility. According to Marcus theory the rate of charge transfer between two identical molecules ω can be calculated as:[58]

$$\omega = \frac{J^2}{\hbar}\sqrt{\frac{\pi}{\lambda k_B T}}\, e^{-\frac{\lambda}{4k_B T}}$$

where λ is the reorganization energy and J is the transfer integral; $\hbar, k_B,$ and T are Planck's constant, Boltzmann's constant, and the temperature. The reorganization energy λ is strongly dependent on the structure of the molecules and decreases as the

size of conjugated core of the molecule increases. The transfer integral *J* describes the probability of an electron hopping between two molecules, which is sensitive to the relative positions and orientations of neighboring molecules. Usually, *J* decays exponentially with increasing intermolecular distance. Hence, materials packing more compactly should exhibit higher charge mobility. *J* is also strongly related to the lateral orientation and configuration of neighboring molecules. Müllen and co-workers[57] computed the transfer integral *J* as a function of the azimuthal rotation angle for several symmetric PAH molecules (Figure I-3). Figure I-3 shows that transfer integral *J* varies dramatically with the orientation angles, which will impact the charge carrier mobilities for the resulting materials. For most PAHs, except for the triphenylene, charge carrier mobility is maximized in the eclipsed configuration of the central core (0° or 60°). Furthermore, the value of *J* decreases sufficiently rapidly that an angle of only 10° from the maximum can half the value of *J* (leading to a 4-fold decrease in charge carrier mobility). To maximize *J*-values and thus improve the performance of PAH materials, it is indispensable to have the capability of tuning the intermolecular orientation of these stacked molecules; this requires a better understanding of the intermolecular interaction between PAH molecules.

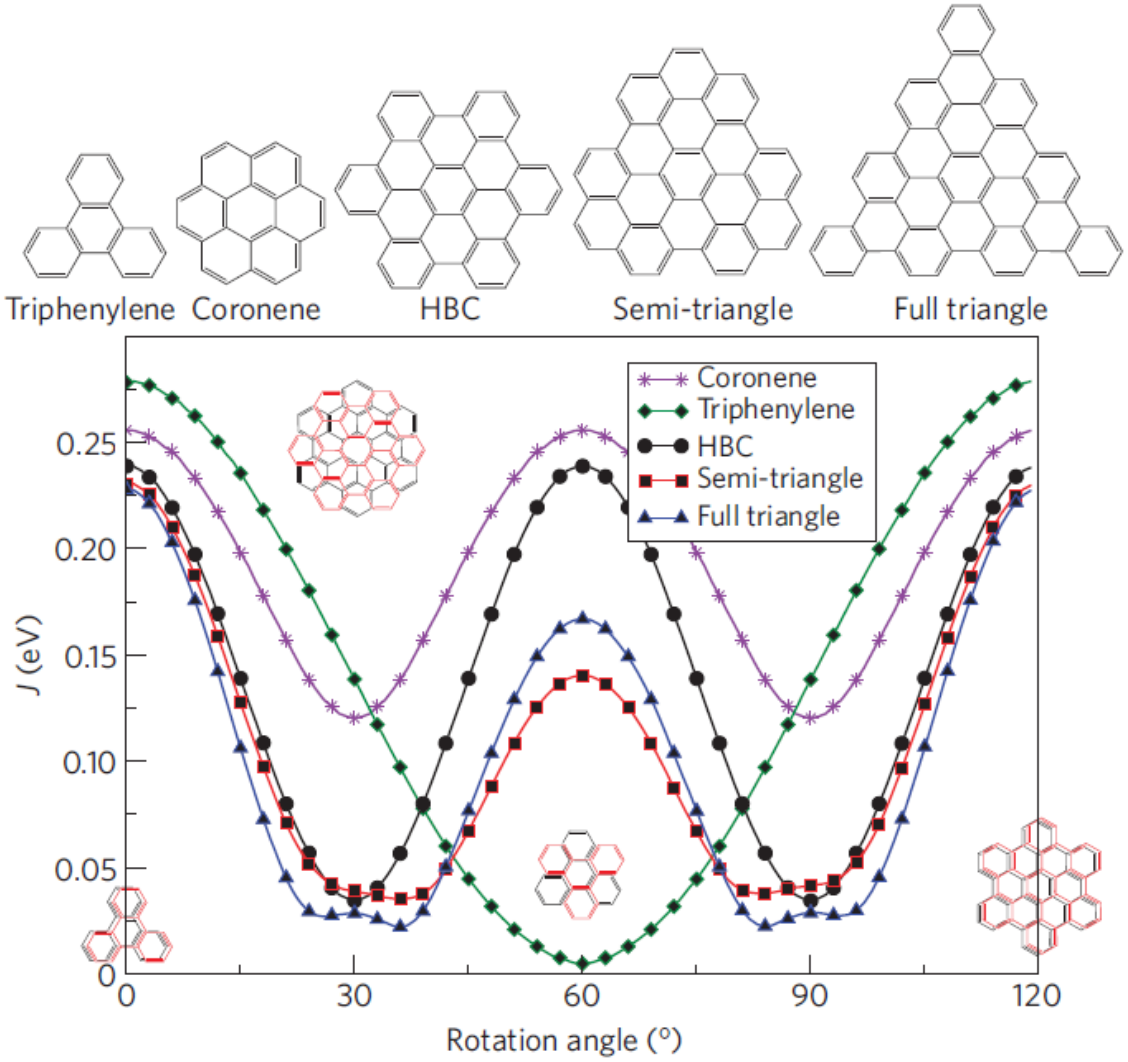


Figure I-3. Absolute value of the transfer integral J as a function of the azimuthal rotation angle for several symmetric polyaromatic hydrocarbon cores. Modified with permission from *Nature Materials* 2009, *8*, 421. Copyright 2009 Nature Publishing Group.

Both experimental and computational studies have been reported regarding the intermolecular interactions and configurations of discotic PAH molecules.[59-64] Many of these studies revealed that molecules inherently stack in a staggered arrangement, which is unfavorable for charge transport. For example, Feng *et al*.[59] synthesized triangle-

shaped PAHs with alkyl chains, which 2D-WAXS showed to pack into helical columns with adjacent molecules exhibiting relative orientations differing by 30°. Wheeler[60] reported computed interaction potential energies as a function of orientation angle for a series of homodimers of substituted coronenes and HBC, finding that the homodimer of unsubstituted HBC favors the eclipsed configuration by up to 5-7 kcal/mol. However, this gap can be overcome through the judicious introduction of complementary pairs of substituents. Andrienko *et al.*[63] performed MD simulation using the OPLS-AA force field for several alkyl-substituted HBC stacking structures. These simulations showed a transition of lateral twist angle toward eclipsed configuration and broader distributions upon converting from solid state to liquid-crystal state for all substituents.

In contrast to planar PAH molecules, our understanding of the molecular packing of curved PAHs is limited. Curved PAHs are forced out of planarity either by embedded nonhexagonal rings[65, 66] or by steric strain from atom crowding.[67] The first studies of the assembly behavior of curved PAHs and attempts to fabricate devices based on curved PAHs started about ten years ago.[49, 67-70] Similar to discotic molecules, curved PAH molecules often pack into columns that then align in a regular lattice.[49, 67] However, the nonplanar structures of curved PAHs yield unique intermolecular contacts within such stacks that are unavailable to planar PAHs. Wu *et al.*[71] reported columnar stacking of a bowl-shaped fragments of C_{70} in which the bowls slip from side-to-side due to the edge-bowl interaction. Intriguingly, all of these bowls were found to align in one direction and thus create a polar crystal.

Various intermolecular contact models of PAHs in materials also give rise to the possibility of improved performance in optoelectronic devices. For example, because the π surfaces of contorted hexabenzocoronenes (c-HBC) in columnar stacks can approach each other closely, such curved systems may be able to achieve higher charge carrier mobilities than those of planar HBC.[67] To design high performed electronic devices using curved PAHs, we must have a sound understanding of the intermolecular interactions exhibited by these molecules.

Curved PAHs can be grouped into three categories: bowl-shaped molecules,[72-75] contorted molecules,[67, 76] and twisted molecules.[77, 78] In the present dissertation, we focus on the former two categories. In Chapter IV, we present our work on understanding local orientations within columnar stacks of a series of bowl-shaped molecules (sumanene, sumaneneone, and sumanenetrione). In Chapter V, we study the intermolecular interactions in homodimers of contorted molecules (*i.e.* saddle-shaped or doubly-concave molecules).

CHAPTER II

AARON: AN AUTOMATED REACTION OPTIMIZER FOR NEW CATALYSTS

2.1 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 catalyst is still far from routine.[28, 35, 79-88] 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 100s of transition state (TS) optimizations required to accurately predict the stereochemical outcome of these reactions.

The selectivity of an asymmetric catalytic reaction stems from the difference in relative rates of formation of stereoisomeric products. In simple systems under Curtin-Hammett control, the difference in rates can be attributed to the difference in free energy between stereocontrolling TS barriers 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. Consequently, the number of accessible TS structures even for a single elementary step in a reaction is often huge.[37, 89] 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, there are a growing number of seemingly simple reactions for which multiple conformations or configurations of a given TS structure or even different elementary steps have been shown to impact the selectivity.[37, 88] 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 substrates and catalysts combinations 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 and 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[90, 91] and synchronous transit-guided quasi-Newton (STQN) methods[92] can often locate the exact TS structure. However, the success of these TS optimization methods relies heavily 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 structure.[93-100] Double-ended methods require structures of both reactant and product to construct a discretized reaction path connecting them, along which TS guesses are generated.[101-105] In terms of rapidly localizing many TS structures, several automated TS

searching and reaction-path exploration approaches have been reported.[44, 106-108] For instance, West *et al*.[107, 108] have developed a high-throughput automated transition state searching method (AutoTST) for high-throughput kinetics. Jacobsen *et al*.[44] recently published details on the AutoTS code implemented in Jaguar,[109] 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, two largely unmet challenges encountered in studies of asymmetric catalysis are the need to sample configurations and conformations. For example, for 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.[85, 110] At the same time, for flexible substrates or catalysts, as well as those with rotatable substituents (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 conventional 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.[85, 87, 111, 112] For example, DiRocco *et al*.[87] 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*.[85] 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. Paton *et al*.[112] demonstrated the utility of sampling conformations using a Monte Carlo Multiple Minimum (MCMM) algorithm using the semi-empirical method PM6-DH2 following by optimizations of low-energy conformers using DFT. 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)[113] 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 100s 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,

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 softwares 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 organocatalytic and transition-metal catalyzed reactions, with a focus on stereoselective transformations.

2.2 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 open-source computational toolkits have been developed by several groups.[114, 115,116] For instance, Kulik and coworkers[115] 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 electronic structure packages at various level of theory.[117] Representative molecular structures created using these classes are depicted in Figure 1.

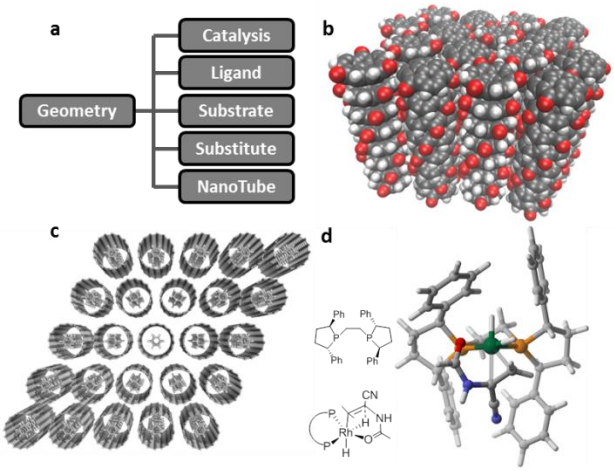


Figure II-1. a) Subclasses derived from the `Geometry` class; b) Initial solid-state structure of columnar stacks of sumanenetrione;[118] c) solid-state structure of aligned [8,8]-carbon nanotubes with included stacked benzenes; d) Transition state structure for a Rh-catalyzed asymmetric hydrogenation reaction.[43]

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 Gaussian09 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*.[43] 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 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:

```
$cata->map_ligand($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 atoms 7 and 8 with methyl groups and the CN group starting with atom 5 with $COOCH_3$. These processes are depicted in Figure 2b, and yield a good guess for the corresponding transitions 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.

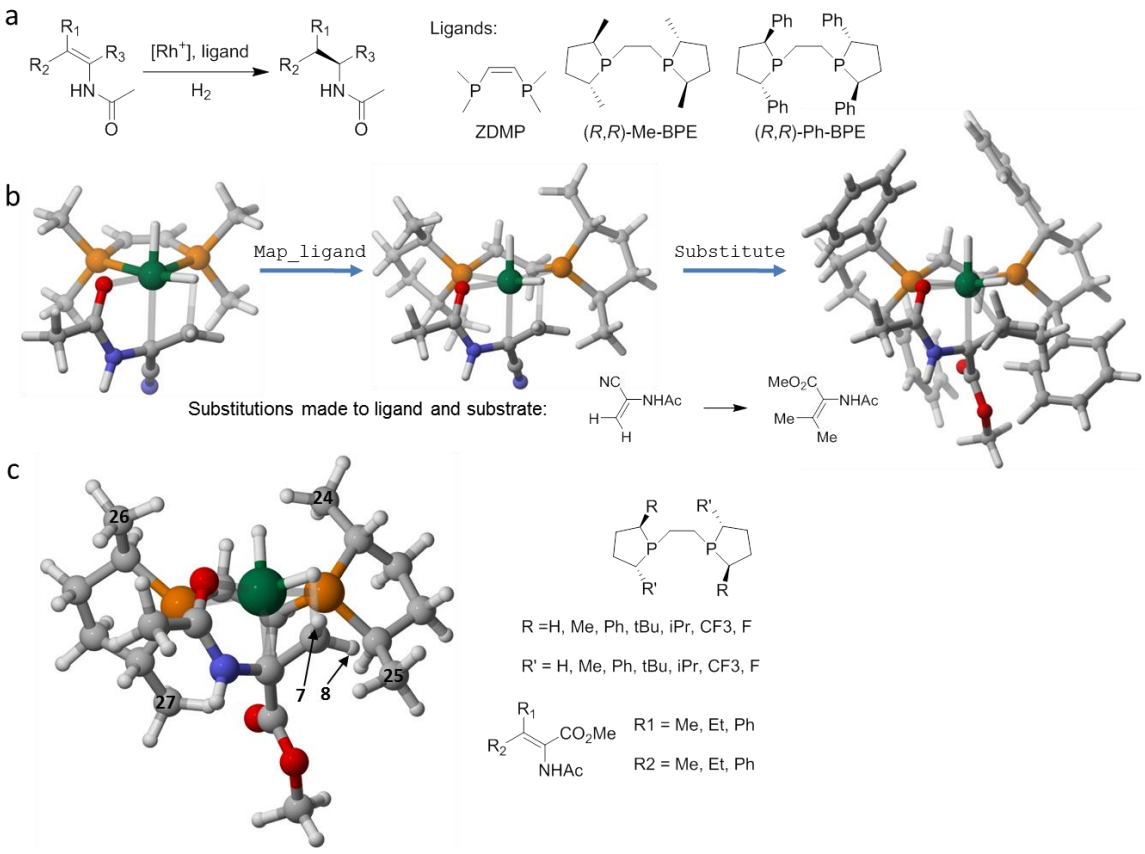


Figure II-2. Construction of initial TS structures for the Rh-catalyzed hydrogenation of enamides using C_2-symmetric phosphoric ligands, starting from a reported TS structure from Wiest *et al.*[43] a) Overall reaction and key ligands; b) replacement of ZDMP with (*R*,*R*)-Me-BPE via map_ligand and 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 non-Perl-based scripts. For instance, the following system call will replace the ZDMP ligand in `ts1.xyz` with (*R*,*R*)-Me-BPE and print the resulting coordinates to `STDOUT`:

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

2.3 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 intermediates structures. Using AARON, users can rapidly compute the 100s 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.

2.3.1 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. Preset levels of theory can be defined either system-wide or by each user, allowing for very simple input files for AARON for routinely-used levels of theory. In the input file, users specify catalysts to be screened by indicating either substitutions on the ligand found in

the TS template library or new ligands. Additional substrates are specified via substitutions of the substrates found in the template library.

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 the initial TS structure for new ligands and substrates based on the structures in the template library;
2. performing a series of constrained and unconstrained geometry optimizations, followed by harmonic vibrational frequencies and potential higher-level single point energies; and
3. analyzing the resulting structures.
4. There are typically many TS structures for each substrate/catalyst combination, and each step (1-5) for each catalyst/substrate combination is run as separate job in a typical HPC environment. This results in high throughput, particularly for clusters with a large number of individual nodes.

To maximize efficiency when varying both catalyst and substrates, AARON performs a hierarchical TS 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.

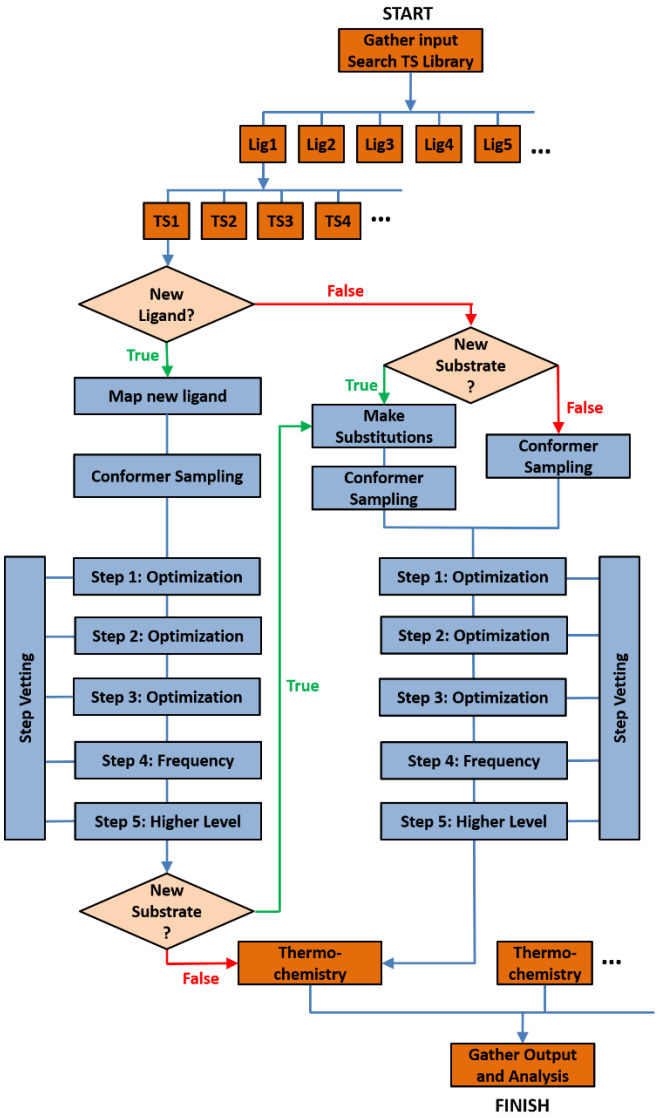


Figure II-3. Overall workflow of AARON.

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 user-defined low level of theory (the default is PM6). This

refines the initial TS structure and removes steric clashes, reducing the number of optimization steps in downstream optimizations at higher levels of theory. In Step 2, an energy minimization is performed at a user-defined DFT level of theory with constraints added to all forming/breaking bonds to yield a structure that should closely resemble the real transition state. In Step 3, AARON performs a full TS optimization using the Berny algorithm.[90, 91] At each point in this process there are many possible issues that can arise, which are caught by the TS vetting process (see Section 5.3.3 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 and can be readily ported to different HPC clusters with popular queueing systems. Moreover, AARON can be stopped and restarted at any point. All 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 and facilitates data organization. A utility is included that automatically constructs tables of absolute energies, enthalpies, and free energies as well as all optimized structures for inclusion in Supporting Information, and capabilities are currently in development that will populate databases with computed TS properties for informatics applications.

2.3.2 Construction of Initial Transition State Structures

Central to locating TS structures is the generation of an initial TS guess for the corresponding catalyst and substrate. AARON uses geometries from previously located TS structures (the TS template library) as templates for new catalysts and substrates. This approach is also one of the core strategies used to generate initial TS structures in AutoTS.[44] 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 TS template 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 and include 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.[119] Starting from these TS structures corresponding to different conformations and/or configurations, AARON detects substituents and

executes a hierarchical search of low-lying conformers (see Section 5.3.4 below for details).

An example of a TS template library can be taken from work from Liu *et al.*,[120] who conducted a thorough study of the regio- and stereoselectivity of Rh-catalyzed carboacylations of benzocyclobuteneones. Liu *et al.*[120] considered six configurations of the stereocontrolling olefin migratory insertion step for this reaction (**TS1**-**6** in Figure II-4), 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 were relatively high-lying for dppb. Using the dppb-based structures already computed by Liu *et al.*[120] as a TS template library, AARON can readily compute all structures leading to both the major and minor stereoisomer for the chiral ligand SEGPHOS.

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. 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 those of 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).

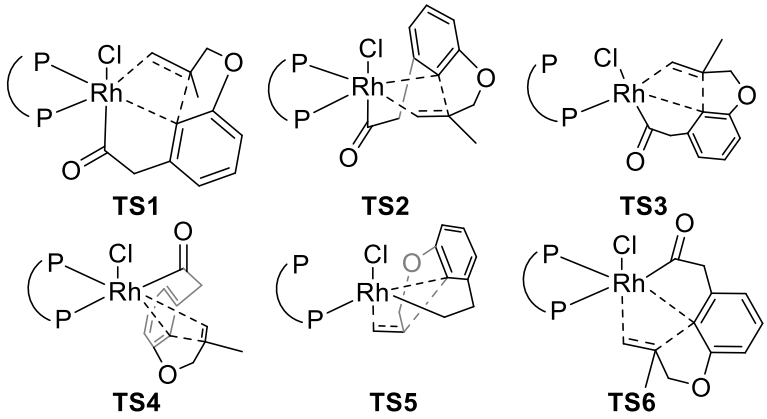


Figure II-4. Example of the multiple configurations and substrate-catalyst orientations that must be included in a TS template library for a Rh-catalyzed carboacylation reaction studied by Liu *et al*.[120]

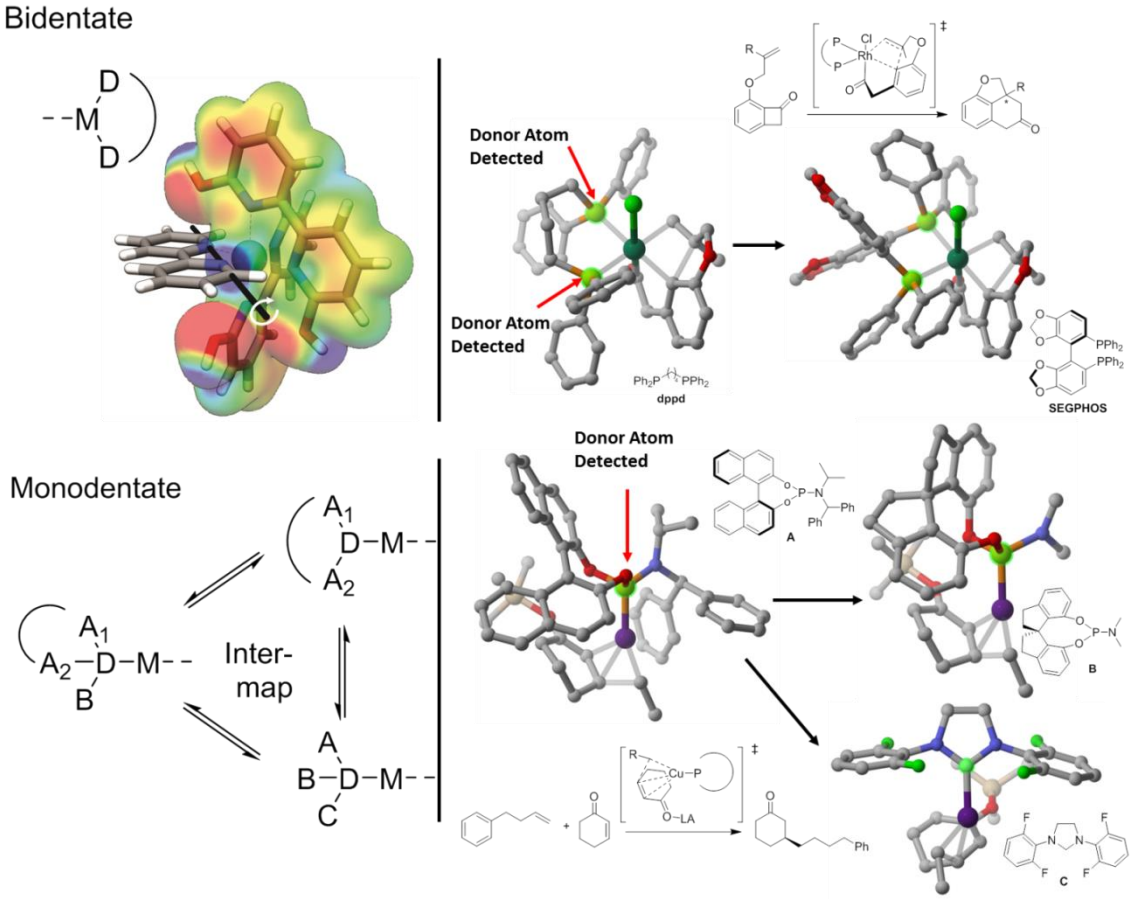


Figure II-5. Examples of replacing bidendate and monodentate ligands: (top) replacing dppb with SEGPHOS in the TS structure computed by Liu *et al.*[120] for a Rh-catalyzed carboacylation reaction; (bottom) replacing a BINOL-based phosphoramidite ligand with both a SPINOL-based analog (B) and an NHC (C) for the Cu-catalyzed conjugate addition studied by Paton *et al.* [35]

For multidentate ligands, AARON maps new ligands by minimizing the root mean-squared deviation (RMSD) between all donor atoms of the new and old ligands. For bidentate ligands, the two donor atoms alone do not uniquely determine the position of ligand. 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 (see Figure 4a). For example, Figure 4a shows the

replacement of the bidentate ligand dppb with SEGPHOS in TS1 from Scheme 1.[120] For monobidentate ligands, the ligand can fall into several classes depending on the presence of cyclic substructures (see Figure 4b). AARON can automatically map among these different classes of cyclic and acyclic monodentate ligands. For example, Figure 4b 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*.[35] with both a SPINOL-based analog (**B**) and an NHC (**C**).

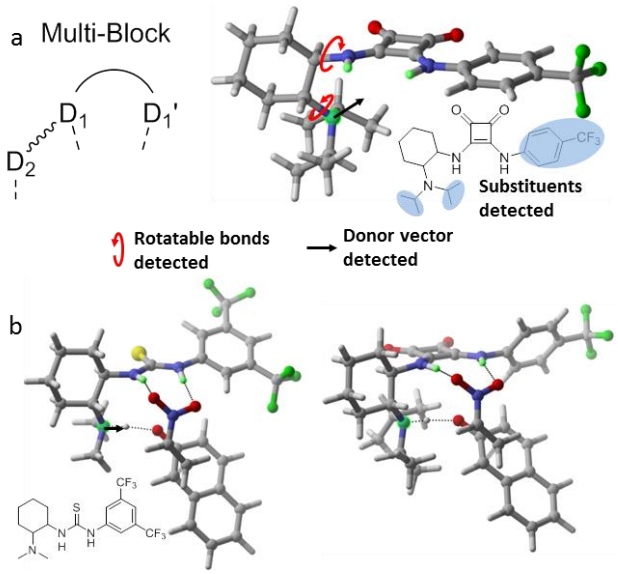


Figure II-6. a) Example of a 'multi-block' catalysts in which multiple atoms of the catalyst connected by flexible linkers bind the substrate. For such systems, AARON automatically detects both central and remote donor atom groups as well as flexible covalent bonds linking these groups. Substituents are also detected automatically. b) Mapping of a squaramide-based catalyst onto a TS structure for a thiourea-catalyzed addition of nitroethene to a β-napthol.

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 automatically divides these donor atoms into those on the core of the catalyst (central donors) and 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 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 of the catalyst (see Figure 5a). For example, for the asymmetric dearomatization reaction catalyzed by a bifunctional hydrogen bonding catalyst presented in Figure 5b,[121] a thiourea-based organocatalyst is replaced by a squaramide-derived catalyst with a distinct backbone and sterically-congested tertiary amine.

2.3.3 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. Some conformational changes must still be manually added to the TS template library, including macrocycle conformations and ring-flips. This represents a considerable time-savings for the user, as searching these conformations manually is tedious. AARON utilizes a rule-based hierarchical conformational search 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.[122] The rules are: 1) conformers are

sampled only for substituents detected by AARON or specified by the user as new substituents; 2) torsional angles for each substituent are determined based on its symmetry; 3) conformations are searched hierarchically such that conformational searching for later substituents begin only after completion of the conformer search for earlier substituents. In contrast to many of the tiered conformational search algorithms[85, 87, 111, 112] 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 and QM potentials. 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 rotamers 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. By sampling conformations according to the symmetry of substituents, we limit the conformational space to be explored. 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 employed. 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.

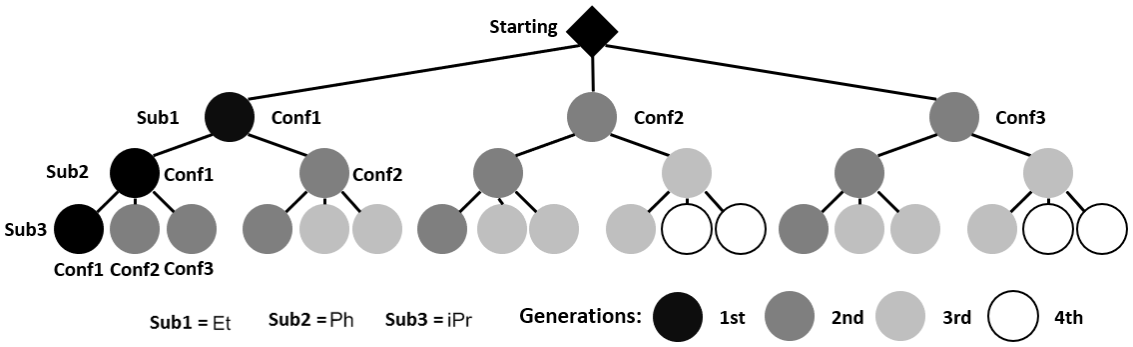


Figure II-7. Example of the hierarchical searching of a fictitious system with three substituents (Et, Ph, and ^{i}Pr).

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, 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.

2.3.4 Thermochemistry

AARON provides predicted stereoselectivities 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.[123] These values can be computed based on the level of theory used for the geometry optimizations and vibrational frequencies or based on 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), we consider a Boltzmann weighting over all unique conformers to compute the effective energy for each TS. For instance,

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

The final predicted stereoselectivities are then based on a 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

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

2.3.5 Step Vetting and Error Checking

Since transition state optimizations are prone to failure, AARON checks all running jobs periodically to ensure 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 fix 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 the 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 general rules derived from our experience locating transition states. 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.

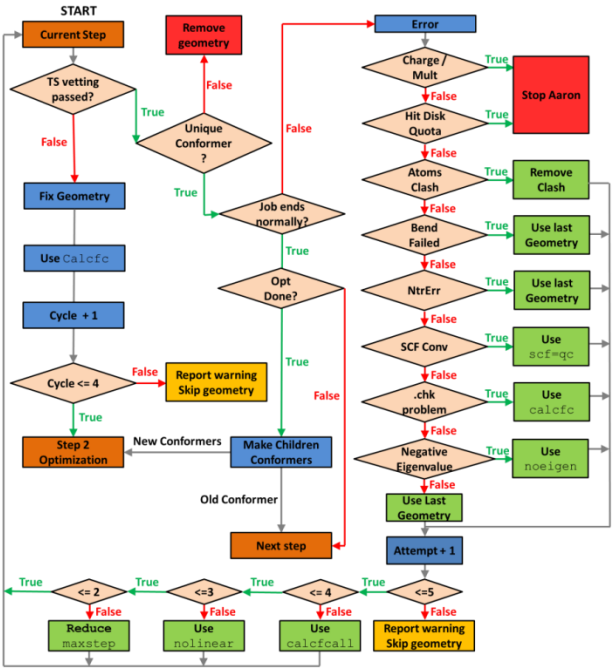


Figure II-8. Procedure for 'Step Vetting', in which AARON constantly monitors all jobs to ensure that correct TS structures are located. For jobs that die with no recognizable error, the last geometry is used and the job restarted.

2.3.6 Locating Single TS Structures

AARON also includes a series of command-line utilities that complement the main AARON workflow. Here, we highlight `TS_search.pl`, which facilitates the location of a single TS structure. This is useful, for example, when initially computing structures to construct a TS template library, but can also be used in any context in which one needs to locate a TS structure. `TS_search.pl` follows a simplified version of the AARON workflow presented above, but only searches a single TS structure rather than searching in parallel for all transition states over multiple ligands and substrates. The most notable difference is that `TS_search.pl` executes the entire workflow within a single HPC job. Starting from even a relatively poor initial TS guess, such as those

quickly built using a graphical molecular builder, `TS_search.pl` reliably obtains an optimized TS structure at a chosen level of theory. For instance, Houk *et al*.[124] recently studied the Rh-catalyzed asymmetric alkynation of trihalomethylketones at the B3LYP-D3/6-31G(d)/LANL2DZ level of theory. Given the rather crude starting guess for the key step in this reaction given in Figure 8, `TS_search.pl` yields the correct TS structure without any user intervention.

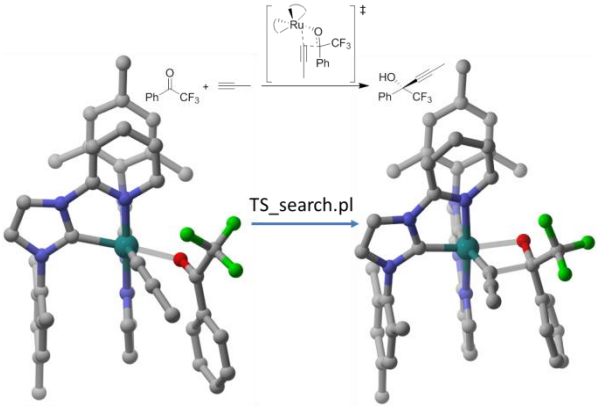


Figure II-9. Structure of an initial TS guess (left) and optimized TS structure provided by the AARON utility TS_search.pl (right) for the Rh-catalyzed alkynylation of trihalomethylketones from Houk *et al*.[124]

2.4 Representative Applications

Below, we demonstrate the power of AARON through applications to various reaction types. In designing of AARON, we strove to generalize the workflow to be as general and flexible as possible, and the following examples show that AARON can be applied to a broad range of reaction types.

2.4.1 Pd-Catalyzed Heck Allenylation

Sigman and co-workers[125-129] have developed a series of Pd-catalyzed redox-relay Heck reactions. Wiest, Sigman, and co-workers[130, 131] have shown computationally that the migratory insertion step in these reactions is stereocontrolling. Most recently, Sigman *et al*.[132] studied the Heck relay reactions of substrates **2a** and **2aa** using the chiral ligand 2-t-Bu-PyrOx (**L1** in Figure 9a) computationally 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 9b) 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. Starting from the eight unique TS structures (ignoring the conformations of the CH_2OH group) reported by Sigman *et al*.[132] we constructed a TS template library by removing the CF_3 and *t*-Bu substituents from the chiral ligand as well as the CH_2OH and CH_3 substituents form the substrate. The cyclohexanone of substrate **1** can potentially have two chair-like 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.*[132] to locate all low-lying TS structures for substrates **2a** and **2aa** by making substitutions on the templates from this TS template library and systematically searching for conformations of these substituents. Systematically considering all orientations of the substituents combined with the two chair-like conformations of the cyclohexanone leads to 192 possible TS structures for each substrate (**2a** and **2aa**). Without any user intervention, AARON located 90 TS structures for **2a** and 94 for **2aa**. The 'missing' TS structures are primarily due to initial conformations that converged to lower-lying conformations during the optimizations.

Relative free energies for the computed TS structures for substrates **2a** and **2aa** are plotted in Figure 9c. 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 9c). 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 9c 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 results highlight the dangers of manually searching for stereodetermining transition states using conventional tools, since one can easily omit 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.

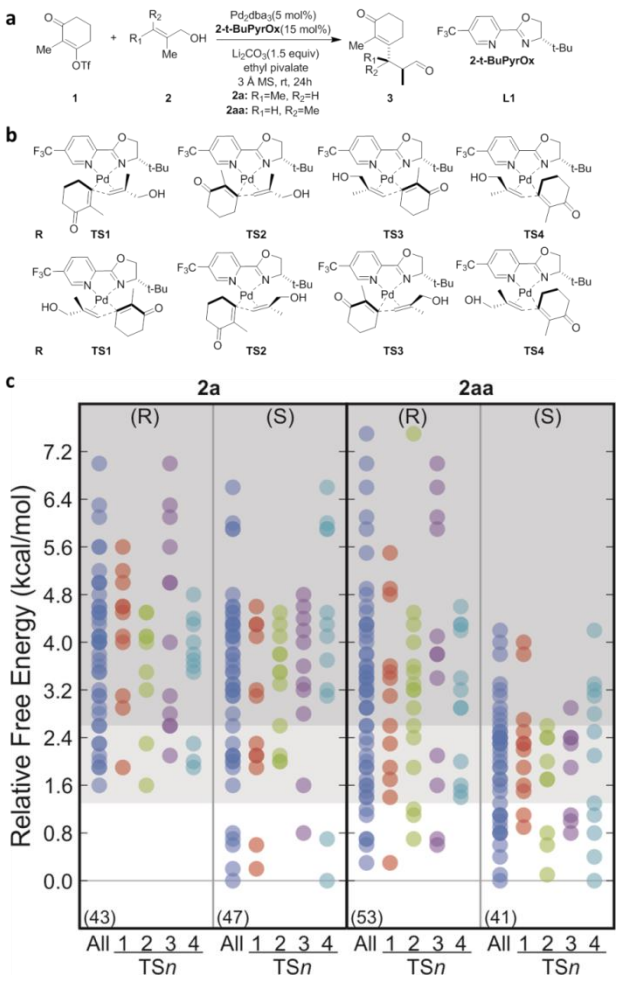


Figure II-10. a) Pd-catalyzed redox-relay Heck reaction from Sigman and co-workers[132] using the chiral ligand 2-t-BuPyrOx. b) Eight TS structure that formed the TS template library for the application of AARON. c) and d) Relative free energies for TS structures located by AARON based on the eight TS structures in the TS template library. The gray shaded regions denote energy ranges of TS structures expected to have Boltzmann populations of <1% (dark gray) and <10% (light gray).

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, with little effort one can screen a given catalyst across 100s 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.

2.4.2 Rh-Catalyzed Hydrogenation of Enamides

Transition-metal-catalyzed asymmetric hydrogenations represent a cornerstone of reactions for building a wide range of optically active compounds,[10] and chiral phosphorus ligands have proved highly effective catalyzing this class of reactions. However, detailed mechanism for this reaction is still unsettled. AARON was previously used to probe the mechanism as well as explain the stereoselectivities of the Rh-catalyzed asymmetric hydrogenation of enamides shown in Figure 10.[88] Previous computational studies[133-136] and our preliminary transition state searches revealed two key steps for this reaction in which hydrides are 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.

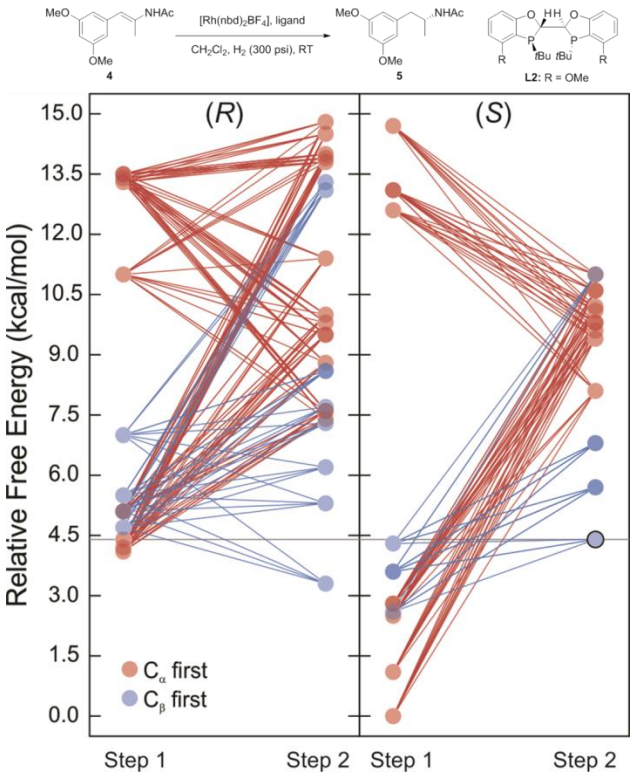


Figure II-11. Relative free energies for TS structures for the asymmetric hydrogenation of enamides for the first and second hydride transfer steps (Steps 1 and 2) following two possible mechanisms (hydride transfer to C_α first or C_β first) for formation of the major (R, left) and minor (S, right) stereoisomeric products. Colored lines indicate pathways between the TS structures for two steps, while the horizontal gray line denotes the energy of the rate-limiting TS structure leading to formation of the minor stereoisomer. Data from Ref. [88].

Using AARON, we screened six ligands and located a total of ~250 TS structures spanning the full 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 10 shows the relative free energies of TS structures for **L2** for the two hydrogenation steps and the two possible mechanisms. There are paths connecting many of these TS structures for the two steps, leading to a complex web of possible reaction paths. For each reaction paths, the step with higher free energy is the rate-limiting step. As seen in Figure 10, for some pathways the first hydride 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. Thus, As seen for the previous example, 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, and the predicted selectivities agree very well with experimental data.[137] 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.

2.4.3 Lewis-Base Promoted Propargylation of Aromatic Aldehydes

Finally, we note that AARON has been successfully applied to organocatalyzed reactions, including bidentate Lewis-base catalyzed allylations and propargylations of

aromatic aldehydes.[138, 139] 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 ligands bound to Si.[140] Each of these five configurations results in a pair of TS structure 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.[37] Lu and co-workers[37, 141] showed accurate stereochemical predictions for these reactions requires the computation of all ten or 20 possible structures, precluding the study of more than a few examples of these reactions if done manually.

Starting from a TS template library taken from Lu *et al*.,[37, 141] Rooks *et al*.[138] 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*.[139] 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*), including 12 catalysts predicted to exhibit stereoselectivities exceeding 95%. The large number of TSs for those catalysts also revealed broad trends in the origin of stereoselectivity in this reaction that would have been difficult to unravel by studying a limited number of instances. Finally based

on these data, Doney *et al.*[139] proposed a novel catalyst predicted to provide very high stereoselectivities by preferentially stabilizing a particular TS structure.

2.5 Summary and Concluding Remarks

The ability to reliably and rapidly predict the stereoselectivities of complex catalytic reactions across different chiral ligands and substrates is a prerequisite for effective computational catalyst design. 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) identify TS structures with precise error control and geometry vetting, 3) search conformers in a parallel and hierarchical way, and 4) predict selectivities through a Boltzmann weighting of multiple TSs 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. Additional features of AARON include the ability to readily screen combinations of ligands and substrates from a simple input file and a number of utilities to facilitate the organizing, storage, and publication of data generated by AARON and to locate individual TS structures.

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 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.[142] There have been a number of heuristic-guided configuration searching methods developed that are applicable to catalytic reactions,[95, 143-148] and methods based on bond-connectivity have been used to explore multiple reaction pathways for transition metal catalysis system.[144,65,67] We can forecast combing AARON with such approaches to automatic reaction network exploration in order construct a more complete tool for computational catalyst design.

Finally, we note that the vast quantities of structural and energetic data generated for TS structures with AARON can facilitate the application of modern informatics tools to catalyst design. For instance, there has been success in applications of multi-variate regressions to experimentally-generated stereoselectivity data as a means of both understanding stereoselectivity and designing better catalysts. AARON opens up the door to such informatics applications based solely on computed data.

CHAPTER III

AUTOMATED QUANTUM MECHANICAL PREDICTIONS OF ENANTIOSELECTIVITY IN A RH-CATALYZED ASYMMETRIC HYDROGENATION[*]

3.1 Introduction

Modern quantum chemistry holds tremendous potential power for the design of chiral ligands for transition metal catalyzed reactions;[81, 86, 149-151] however, the routine computational design of new chiral catalysts remains a significant unmet challenge.[152, 153] This is due to a number of factors, including the breadth of possible mechanistic pathways for a given transformation combined with the need to accurately compute relative reaction rates accounting for myriad subtle effects (solvent, additives, *etc*.). Even for mechanistically simple reactions, reliable quantum mechanical predictions of stereoselectivities often require the optimisation of 100s of molecular structures, including many transition states (TSs), due to the possibility of multiple stereocontrolling TSs and the presence of many thermodynamicaly accessible conformations.[154] Such computations are tedious and time consuming, and the experimental synthesis and testing of new chiral ligands remains the preferred route to their development.

---

[*]Adapted with permission from "Automated Quantum Mechanical Predictions of Enantioselectivity in a Rh-Catalyzed Asymmetric Hydrogenation" by Y. Guan and S. E. Wheeler, 2017. *Angew. Chem. Int. Ed.* **56**, 9101. Copyright 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

Here, we describe a new version of our computational toolkit AARON (An Automated Reaction Optimizer for New catalysts)[33, 155-157] that provides automated quantum mechanical predictions of stereoselectivities of transition metal catalyzed reactions with well-established mechanisms[158] but multiple stereocontrolling transition states. This new version of AARON is used to screen C_2-symmetric phosphorus ligands for the asymmetric hydrogenation of (*E*)-β-aryl-*N*-acetyl enamides (Scheme 1), providing selectivity data in good agreement with experiment.[137] Data is also presented for a new ligand that is predicted to provide enhanced stereoselectivity for this reaction.

Transition-metal catalyzed asymmetric hydrogenations provide efficient routes to a wide range of optically active compounds,[10] and chiral phosphorus ligands have proved highly effective in such transformations.[159-165] Tang *et al*.[137] introduced a strategy for the development of C_2-symmetric phosphorus ligands based on the creation of a deep chiral pocket, as exemplified by WingPhos (**L5**, Scheme 1).[137] In contrast to popular ligands such as BINAP[160] and Josiphos,[166] WingPhos provided excellent stereoselectivites in Rh-catalyzed asymmetric hydrogenations of (*E*)-β-aryl-*N*-acetyl enamides, providing access to enantiopure chiral β-arylamines (see Table 1).[167-169] WingPhos has also proved effective in other asymmetric transformations, supporting the merits of designing ligands with deep chiral pockets.[170, 171]

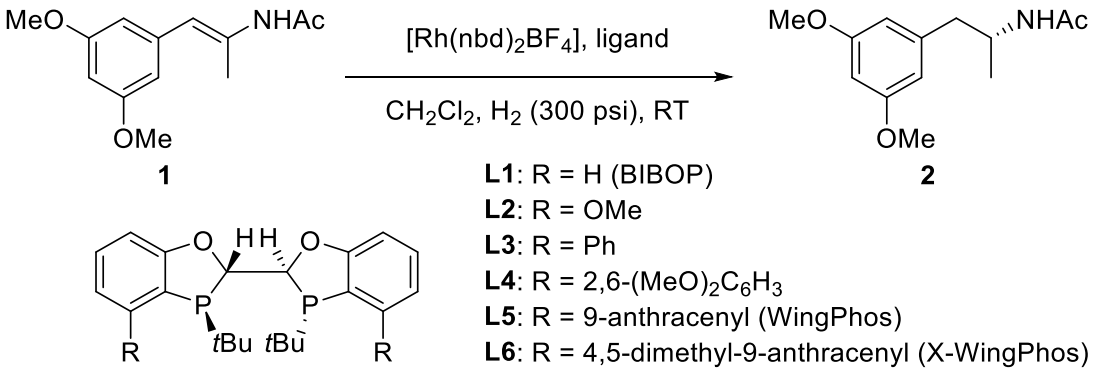


Figure III-1. Rh-catalyzed asymmetric hydrogenation of (*E*)-β-aryl-*N*-acetyl enamides from Tang *et al*.[137]

Table III-1. Experimental and theoretical relative activation free energies ($\Delta\Delta G^{\ddagger}$) and *ee*'s. The primary stereodetermining TS structures for formation of the (*R*) and (*S*) products are also listed (see Scheme 2).[a]

	Exp[137]		Theor[b]			
Ligand	ee	ΔΔG‡	ee	ΔΔG‡	TS(R)	TS(S)
L1	27	0.3	-41	-0.5	TS1α1	TS2α1
L2	11	0.1	11	0.1	TS1β1	TS2β2
L3	35	0.4	58	0.8	TS2α1	TS2β2
L4	74	1.2	78	1.3	TS2α1	TS1β1
L5	97	2.5	>99	6.2	TS1α1,TS2α1	TS1β1
L6			>99	9.5	TS1α1	TS1β1

[a] Positive ee values correspond to excess (R) product, whereas negative values signify excess (S).
[b] Boltzmann-weighted free energy barrier of all (S) pathways relative to all (R) pathways and corresponding ee.

3.2 Computational Section

Computations were performed using Gaussian09[63] paired with AARON.[155]

Geometries were optimized using M06-L with the LANL2DZ basis set with ECP on Rh

and 6-31+G(*d*,*p*) on all other atoms.[172, 173] This combination of basis sets has been shown to provide reliable predictions at a relatively low computational cost.[174-176] The final free energies were based on M06 single point energies using the SDD basis set with ECP for Rh and 6-311++G(*d*,*p*) for all other atoms.[177-179] Free energy corrections were computed within the rigid-rotor harmonic-oscillator (RRHO) approximation. Predicted $\Delta\Delta G^{\ddagger}$ values and enantioselectivities were based on a Boltzmann weighting of all pathways leading to the major and minor stereoisomers of **2** (see APPENDIX B). The M06-L computations utilized density fitting techniques and all computations used the IEF-PCM solvent model (dichloromethane).[180, 181]

3.3 Results and Discussion

Despite advances in the development of catalysts for asymmetric hydrogenations of enamides,[137, 182, 183] the mechanism of these reactions is still unsettled.[43, 133, 134, 184-192] Common among the proposed mechanisms is the importance of the two hydride transfers from the metal to the alkenyl component of the substrate; whether the α- or β-carbon is hydrogenated first varies with both the substrate and catalyst.[133-136] Below, we focus on these two hydride transfer steps.

Rh-catalyzed asymmetric hydrogenations of enamides have been widely studied computationally.[38, 43, 133-136, 186-188, 193, 194] For example, Wiest, *et al*.[43] demonstrated the utility of Q2MM in the prediction of stereoselectivities for these reactions. While such force-field based methods[195-198] provide a powerful tool for the computational screening of potential new ligands, they depend on the presence of a single stereocontrolling TS

that is conserved across all ligands screened. For the reaction in Scheme 1, there are four possible TS structures for the first (TS1) and second (TS2) hydride transfer for both of the enantiomeric products. These differ in the arrangement of the various ligands around the Rh and whether the hydride is transferred to the α- or β-carbon first (see Scheme 2, in which TSXαY and TSXβY refer to TS structures in which the hydride is transferred to the α- and β-carbon first, respectively).[43] Transfer of the first hydride to the substrate yields a hydride complex that can undergo one of two rearrangements, leading to two distinct transition states for the second H-transfer. For example, rearrangement of the hydride complex following TS1β1 can lead to either TS2β1 or TS2β2. It is important to note that rearrangement of the hydride complex following TS1β2 leads to the same possible structures for TS2 (*i.e.* TS2β1 and TS2β2).[199200]

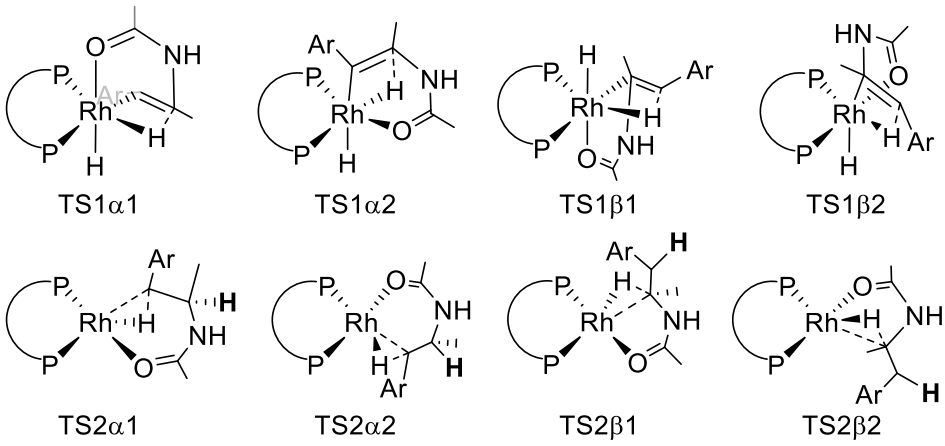


Figure III-2. Classification of the possible TS structures for the first (TS1) and second (TS2) hydride transfers in the reaction in Scheme 1.[161] For TS2, the previously added hydrogen is shown in bold.

As shown below, for the reaction in Scheme 1 either TS1 or TS2 (or both) can impact the stereoselectivity. Consequently, to predict the stereochechemical outcome of

this reaction, at least eight transition states (four for TS1 and four for TS2) should be considered for each enantiomeric product (*i.e.* 16 transition states total). Once the conformational flexibility of the substrate and catalyst are considered (*e.g.* the two rotatable OMe groups on the aryl ring), there can be 100s of distinct TS structures for each ligand.

Free energies for the eight TS structures in Scheme 2 were computed for the reaction in Scheme 1 catalyzed by **L1** for both the major and minor enantiomeric products (with the two OMe groups removed from the substrate). All but two of the TS structures (TS1α2 and TS1β2) lie within 4 kcal mol^{-1} of the lowest-lying TS structure, TS2α1(*S*). The large relative free energies for TS1α2 and TS1β2, which exceed 10 kcal mol^{-1}, derive from steric interactions between the catalyst and the β- aryl group of the substrate. This steric clash is even more severe for the ligands with larger R groups (*i.e.* **L2-L5**) and we were able to discard TS1α2 and TS1β2 from further consideration. Even so, the presence of two OMe groups on the substrate leads to four possible conformations for each of these six remaining TS structures for formation of both the major and minor stereoisomers (48 structures total); these were located automatically using AARON (*vide infra*). Low-lying TS structures are plotted in Figure 1a. In total, there are 32 distinct pathways and 46 TS structures [24 leading to (*S*) and 22 leading to (*R*)]. All lie within 5 kcal mol^{-1}. Moreover, both TS1 and TS2 impact the stereoselectivity, with the first step primarily rate limiting for (*R*) and the second primarily limiting for (*S*). In both cases, the most favourable pathway involves a hydride transfer to the α-carbon first, as proposed by Tang *et al.*[137] based on work from Gridnev

and Imamoto.[187, 189] Given the importance of multiple low-lying TS structures, modelling the stereoselectivity of this reaction requires a Boltzmann weighting over all accessible pathways (see APPENDIX B). Performing such a weighting correctly predicts the low selectivity of **L1** (see Table 1).

For **L2**-L5, there are up to 192 potentially distinct TS structures for each ligand.[201] Systematically optimising all of these structures by hand would be tedious; however, the automation provided by AARON makes such computations not only possible, but practical. AARON automates the optimization of all relevant structures by starting with a library of TS structures for a representative model ligand.[202] AARON then maps key atoms of the new ligand onto the corresponding atoms from structures in the TS library (in this case, the two phosphorus atoms) and then performs a prescribed series of constrained and unconstrained geometry optimizations, as described previously.[156, 157] For each structure, the conformations of simple rotatable groups (*e.g.* OMe) are systematically scanned automatically. Checks performed at each stage of the optimizations ensure that the correct structures are located and duplicate conformations are eliminated.[203] To reduce the chance that key structures are omitted, any new structures identified for a given ligand can be easily added to the TS library and the analogous structures automatically located for the other ligands.

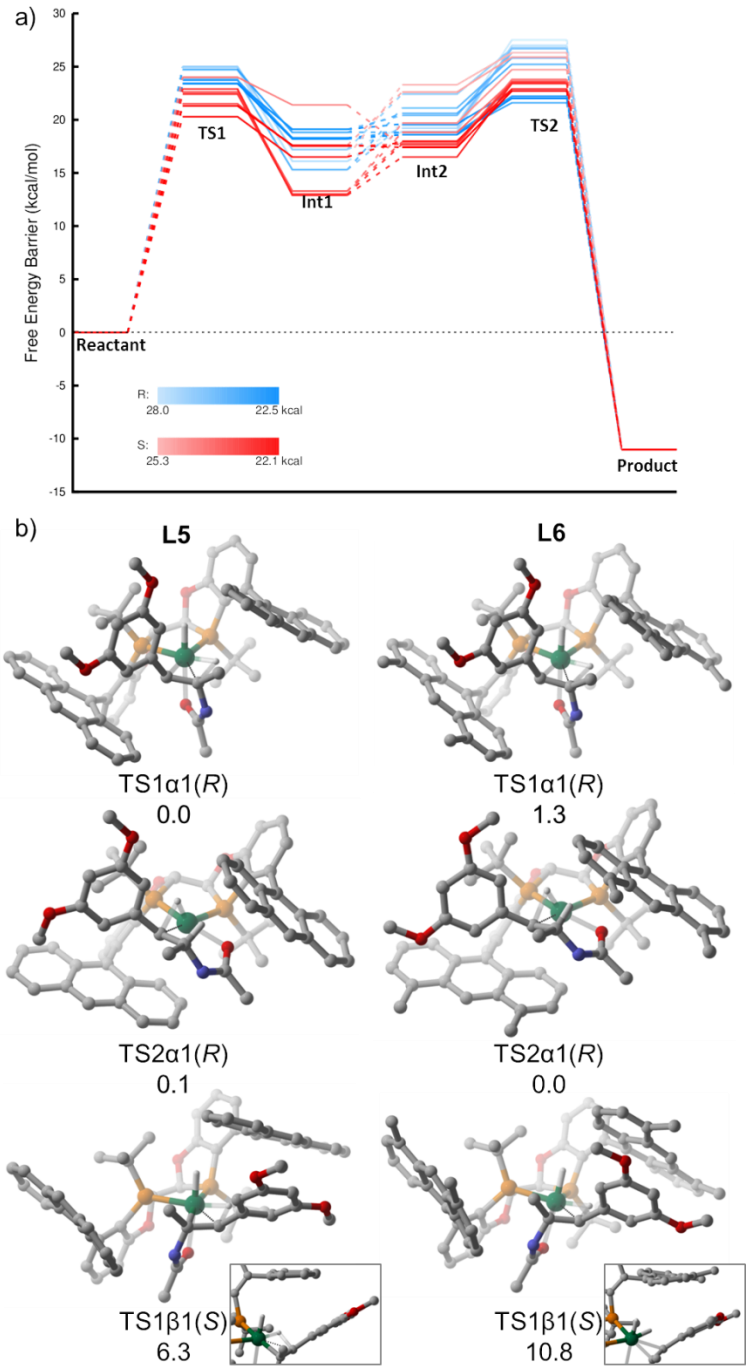


Figure III-3. a) Free energies (relative to separated reactant and catalyst)[204] for thermodynamically accessible hydride transfer TS structures and associated intermediates for the reaction in Scheme 1 catalyzed by L1. b) Key TS structures for L5 and L6. For TS1β1(*S*), the insets highlight the steric clash in the TS leading to the minor product. Hydrogens have been removed for clarity.

Using AARON, we have systematically optimised all viable TS structures for **L2**-**L5**. A Boltzmann weighting over the possible reaction pathways leads to computed *ee*'s in good agreement with experiment (see Table 1).[137] The overall reaction free energy profiles vary from **L1** to **L5**, highlighting the importance of considering multiple possible TS structures for these reactions. For instance, for **L2** the most favourable reaction path involves the hydride transfer to the β-carbon first for both the (*R*) and (*S*) isomers, in contrast to **L1**. For **L3**-**L5**, on the other hand, the dominant pathways leading to the two stereoisomers diverge. That is, for these ligands the formation of the (*R*) enantiomer is dominated by the hydride transfer to the α-carbon first, while formation of (*S*) predominately involves a hydride transfer to the β-carbon first. This is inconsistent with the mechanism presented by Tang *et al.*[137]

For **L5**, AARON predicts an *ee* exceeding 99%, which is consistent with experiment (see Table 1).[137] The primary stereocontrolling TS structures for **L5** are shown in Figure 1b. In contrast to the other ligands, in the case of **L5** the lowest-lying structures for the first and second hydride transfers leading to the (*R*) enantiomer [TS1α1(*R*) and TS2α1(*R*), respectively] are essentially isoergonic. For the most favourable reaction path leading to the (*S*) enantiomer, on the other hand, the barrier for the first hydride transfer is 4.1 kcal mol^{-1} higher in free energy than the second hydride transfer and barrier height for this first hydride transfer is the dominant stereodetermining factor. The most favourable configuration for this step is TS1β1(*S*), which lies 6.3 kcal mol^{-1} higher in free energy than TS1α1(*R*). This significant energy

difference, which underlies the selectivity of **L5**, can be attributed to steric interactions between the aryl groups of the catalyst and substrate (see Figure 1b).

The agreement between the computed and experimental *ee*'s for **L1**-**L5** validates the choice of level of theory and also provides confidence regarding the prediction of stereoselectivities for new ligands. Inspired by this, we extended **L5** to provide greater differentiation between the primary stereocontrolling TS structures. This resulted in **L6**, in which the 9-anthracenyl substituents bear methyl groups at the 4- and 5-positions. As seen in Figure 1b, **L6** leads to a more sterically-demanding binding pocket, resulting in significant distortion of TS1β1(*S*). The result is an increase in free energy of TS1β1(*S*) relative to TS2α1(*R*), as well as a modest increase in the free energy of TS1α1(*R*) relative to TS2α1(*R*), compared to **L5**. The net effect is an increase in the Boltzmann-weighted relative free energy barrier for formation of the minor stereoisomer. In other words, even though **L5** is highly stereoselective,[137] we predict that **L6** will perform even better. Moreover, the predicted hydride transfer free energy barrier is 0.3 kcal mol^{-1} lower for **L6** than that for **L5**, suggesting slightly enhanced catalytic activity.

3.4 Conclusions

In conclusion, we have described a new version of our computational toolkit AARON[155] that is applicable to transition-metal catalyzed reactions. This new version of AARON automates the optimisation and Boltzmann weighting of the many transition states needed to predict stereoselectivities of reactions in which multiple transition states contribute to stereoselectivity, including automated consideration of low-lying

conformations. While such computations could potentially be carried out by hand using conventional tools, the automation provided by AARON greatly accelerates the process and thereby opens the door for the computational screening of chiral ligands for select transition-metal catalyzed reactions. The utility of AARON was demonstrated by screening chiral ligands for the asymmetric hydrogenation of (E)-β-aryl-*N*-acetyl enamides, providing computed stereoselectivites in good agreement with experimental data.[137] These computational data reveal that the major and minor enantiomeric products of this reaction form via two distinct reaction mechanisms. Based on these data, an extended version of WingPhos was designed (**L6**) that is predicted to provide enhanced stereoselectivity and catalytic activity. Although the age of the purely computational catalyst design is not yet upon us,[152, 153, 205] this work represents a key step toward this goal by providing tools for the automated quantum mechanical prediction of stereoselectivities of reactions with well-established mechanisms but variability regarding the stereodetermining step.

CHAPTER IV

INTERCOLUMNAR INTERACTIONS CONTROL THE LOCAL ORIENTATIONS WITHIN COLUMNAR STACKS OF SUMANENE AND SUMANENE DERIVATIVES*

4.1 Introduction

The buckybowl sumanene and its derivatives (Scheme 1) are a prototypical class of bowl-shaped aromatic hydrocarbons that form well-ordered concave-convex columnar stacks in the solid state (Figure 1a).[66, 206, 207] This columnar packing imparts characteristic electronic properties to the resulting materials, including high electron conductivity.[208]-[209] As with discotic liquid crystalline materials,[210] charge-carrier mobilities and other electronic properties of materials comprising bowl-shaped molecules depend strongly on the precise solid-state packing and distributions of local orientations within columnar stacks.[211-213] In the case of sumanene derivatives, local orientations are strongly dependent on substituents. For instance, sumanene (1),[206] sumaneneone (2),[214] and many other substituted derivatives,[215, 216] adopt staggered configurations within each column in the solid state. That is, adjacent stacked bowls are twisted about their central axis by ~60° (e.g. see Figure 1b). This favorability of staggered configurations over eclipsed (Figure 1c) for molecules such as sumanene has been attributed to three complementary effects:[216] relief of repulsive steric interactions

*Adapted with permission from "Intercolumnar Interactions Control the Local Orientations within Columnar Stacks of Sumanene and Sumanene Derivatives", by Y. Guan and S. E. Wheeler, 2017. *J. Phys. Chem. C* **121**, 8541 Copyright 2017 American Chemical Society.

between endo-H and exo-H atoms; favorable CH/π interactions[217] between endo-H atoms of the five-membered rings and the six-membered rings on the other molecule (Figure 1d); and expected electrostatic favorability of the staggered configuration as predicted by analyses of electrostatic potential maps (Figure 1e).

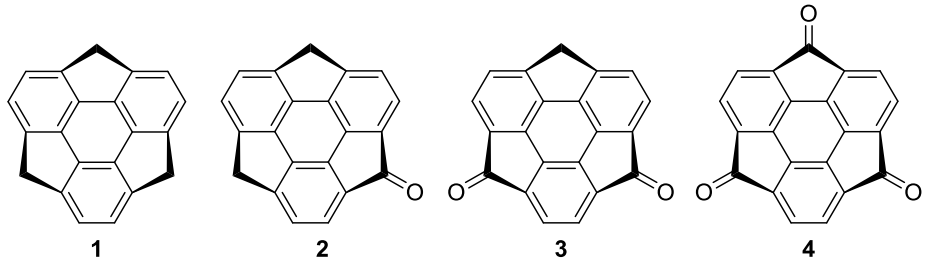


Figure IV-1. Strctures of sumanene (1) and sumanene derivatives (2 - 4).

However, the eclipsed orientations observed in a recently published crystal structure of sumanenetrione (4) have proved more difficult to explain.[218] In particular, analyses of the molecular electrostatic potential (ESP) of 4 suggest a strong electrostatic drive to adopt a staggered configuration. Sakurai et al.[218] explained the eclipsed columnar stacking of 4 based on favorable intercolumnar CH…O interactions, which apparently overwhelm the unfavorable electrostatic interactions within each stack. Recently, Chen et al.[219] studied stacked dimers of corannulene, sumanene, and various substituted derivatives (including 1, 2, and 4) to assess the impact of substituents on charge transport properties. They identified several local minima on the associated stacking potential energy surfaces. Notably, they found that the global minimum for stacked dimers of 4 exhibited a staggered configuration, in contrast to the established

solid-state packing;[218] the energy minimum corresponding to the eclipsed configuration was predicted to be 0.6 kcal mol-1 higher in energy.

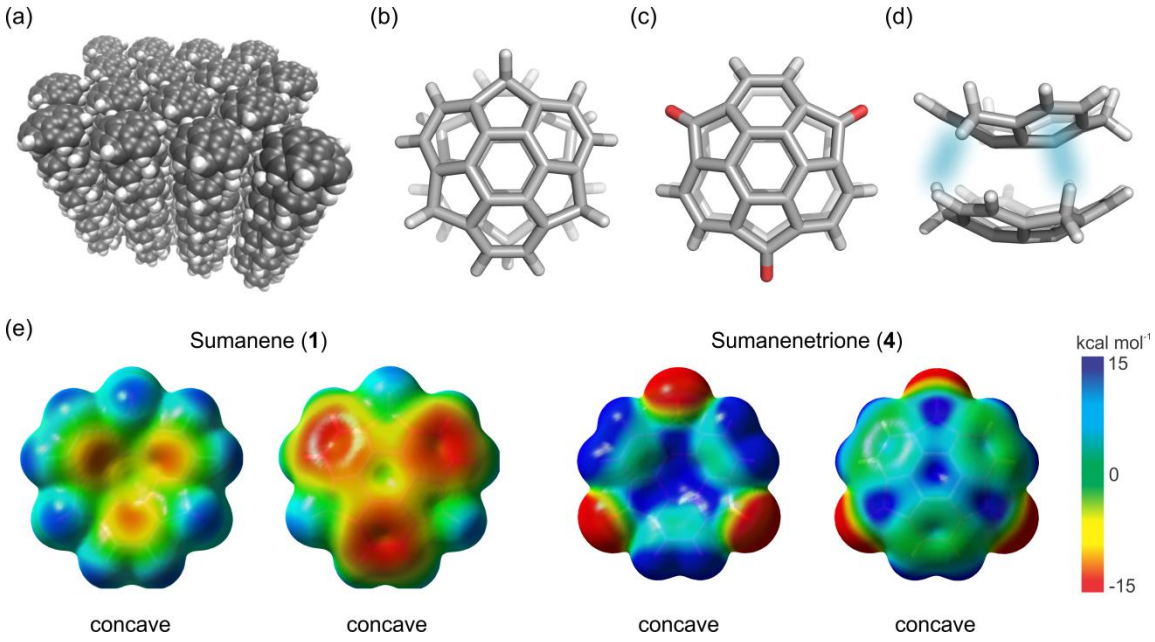


Figure IV-2. (a) Columnar stacked solid-state structure of sumanene; (b) preferred staggered configuration of **1**; (c) preferred cclipsed stacking configuration of **4**; (d) CH···π interactions (highlighted in blue) in the staggered stacked sumanene dimer; (e) ESPs of the concave and convex faces of **1** and **4**.

Recent work from Risko et al.[220] has demonstrated the power of computational chemistry to understand the crystal packing of planar polycyclic aromatic hydrocarbons, building on the vast body of knowledge of stacking interactions involving planar systems.[221-225] However, stacking interactions involving curved polycyclic aromatic hydrocarbons are less well understood,[226] hindering the rational design of bowl-shaped molecules with precisely controlled packing in the solid state. We have examined the relative orientations of 1 - 4 in the gas-phase using dispersion-corrected density functional theory (DFT) and symmetry adapted perturbation theory (SAPT) as well as in

the solid state using classical molecular-mechanics based molecular dynamics simulations, shedding light on the factors that control the packing of these stacked bowl-shaped molecules.

4.2 Theoretical Methods

The structures of monomers and stacked dimers of sumanene **1**, sumaneneone **2**, sumanenedione **3** and sumanenetrione **4** were optimized at the ωB97X-D/TZV(2d,2p) level of theory.[227, 228] The resulting structures were confirmed to be energy minima based on the absence of imaginary vibrational frequencies. Rigid-monomer interaction energies of model stacked dimers were computed as a function of the orientation angle at the B97D/TZV(2d,2p) level of theory. For each angle the optimal bowl-to-bowl distance was determined by computing single point energies for distances ranging from 3.5 to 4.0 Å at intervals of 0.1 Å and fitting a 5th order polynomial to the resulting data (see Appendix B Figure B-1).

We also examined stacked dimers of **1** – **4** using symmetry-adapted perturbation theory (SAPT) at the SAPT0/jun-cc-pVDZ level of theory. This level of SAPT, combined with the jun-cc-pVDZ basis set of Papajak and Truhlar,[229] has been shown to provide excellent predictions of stacking interactions at a modest computational cost.[230] The interaction energy between dimers was decomposed into electrostatic (Elec), exchange repulsive (Exch), induction (Ind) and dispersion (Disp) interactions. SAPT0 computations were performed for stacked dimers every 10° from 0° to 60° or 180°,

depending on the symmetry of the molecule, at the bowl-to-bowl distance determined using DFT.

Finally, the solid state structures of **1** – **4** were simulated by a united atom approach.[61] Carbon atoms were assigned to four main atom types, CA, CM, CO and CB, and further sub-divided depending on the type and positions of peripheral substituents (see Appendix B for more details). The bonded and Lennar-Jones parameters were taken from the AMBER and OPLS force fields,[231-234] which has been demonstrated to be effective in describing stacked columns of discotic molecules.[61, 210] Partial charges were computed using the Merz-Kollman (MK) method[235, 236] at the B3LYP/6-311+G(d,p) level of theory.[237-239] The electrostatic potential fitting involved eight layers and six points per unit area (see Appendix B). For these simulations, we considered systems of 128 molecules arranged in 16 stacks of 8 molecules each. The bowl-shaped molecules were stacked in unidirectional columns for sumanene **1** while in opposite ones for **2**, **3**, and **4** (see Figure B-2), in accordance with available experimental data. The distances between the columns and molecules within columns were increased slightly in the initial structures to avoid clashes. Initial orientation angles between adjacent molecules were 30°, to avoid biasing the systems toward either eclipsed or staggered configurations. After energy minimization, a short equilibrating run was performed at 200 K and 1.0 bar for 4 ns. The Berendsen[240] method was employed for the temperature and pressure coupling model. After the equilibrating process, the production run was performed at 300 K and 1.0 bar for 100 ns to give the final structures. During the production run, Nose-Hoover[241, 242] thermostat and Parrinello-Rahman[243] method were employed for the

temperature and pressure coupling methods, respectively (see Appendix B). To evaluate the MD method, obtained structures were compared with the X-ray crystal structures, focusing on intermolecular distances, bowl depths, and averaged orientation angles between adjacent molecules with respect to the columnar axis. To collect these data, the production run was extended by 1 ns, and 500 snapshots from this 1 ns run analyzed.

Finally, we considered a cluster model of eleven molecules with a central trimer and eight surrounding monomers from peripheral columns, which were extracted from the last snapshot of the MD simulations. Then each monomer was replaced with the DFT-optimized structure, keeping the center of mass and relative orientations of each monomer the same as the MD simulation. The central molecule was then rotated with an interval of 3° to evaluate the interaction energy at the B97D/TZV(2d,2p) level of theory (see AAPENDIX C Figure C-4).

The DFT and MK partial charge computations were done using Gaussian09,[40] while Psi4[244] was used for the SAPT0 computations. The MD simulations were run using GROMACS 4.6.5.[245]

4.3 Results and Discussion

First, we considered the interaction energy, relative to the eclipsed configuration, as a function of orientation angle for model stacked dimers of 1 – 4 (Figure 2a) at the B97D/TZV(2d,2p)[228, 246] level of theory using Gaussian 09.[40] In these one-dimensional scans, the monomers were held rigid and the distance between stacked bowls was optimized for each angle (see Figure 2b). For all four systems, the local minimum corresponding to the staggered configuration is favored over the eclipsed configuration;

however, there is a gradual equalization of the staggered and eclipsed conformations as more carbonyl groups are added. For instance, whereas the staggered configuration of 1 is favored by 2.2 kcal mol-1 over the eclipsed configuration, this energy difference shrinks to 0.6 kcal mol-1 for 4. Intriguingly, the barrier for conversion of the staggered orientation to the eclipsed orientation is nearly constant (2.3 ± 0.03 kcal mol-1). On the other hand, the reverse barrier is quite sensitive to the number of carbonyl groups, growing from only 0.2 for 1 to 1.7 kcal mol-1 for 4.

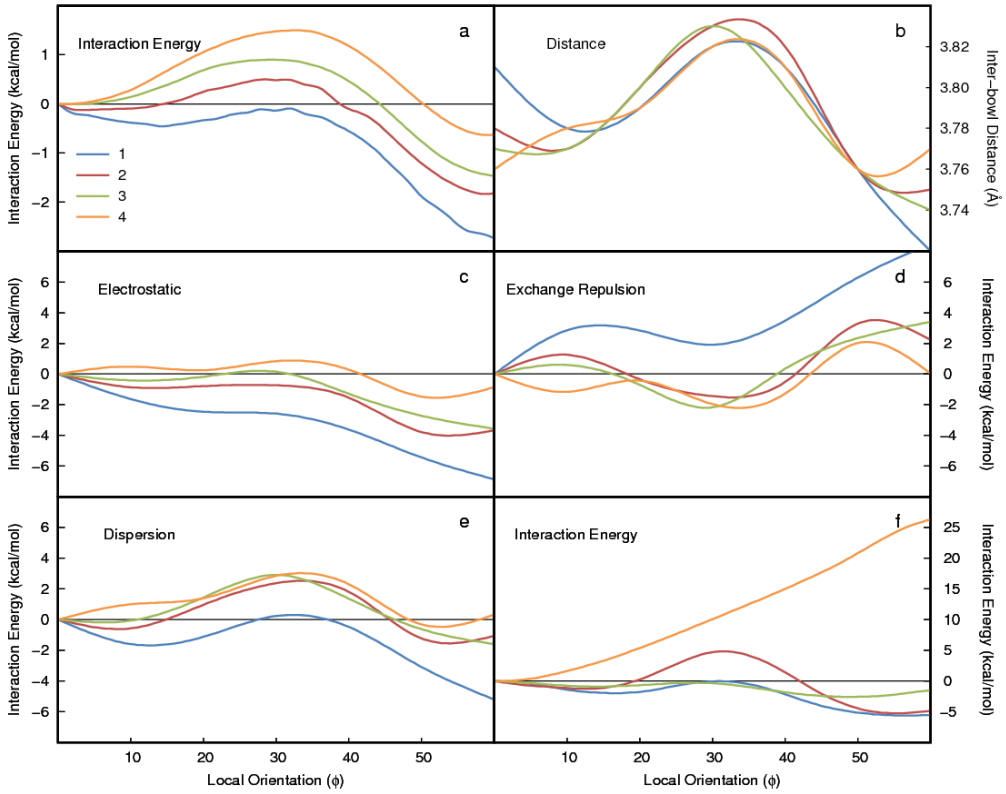


Figure IV-3. (a) B97D/TZV(2*d*,2*p*) energy of stacked dimers of 1 – 4 as a function of the local orientation angle (ϕ), relative to the staggered configuration (ϕ = 0); (b) bowl-bowl distances (*h*, in Angstroms) for stacked dimers of 1 – 4. (c)-(e) SAPT energy components of stacked dimers of 1-4 as a function of the local orientation angle (ϕ), relative to the staggered configuration (ϕ = 0). (f) B97D/TZV(2d,2p) energy of molecular clusters of 1 – 4 as a function of the local orientation angle (ϕ) (for more details, see Appendix B Figure B-4).

To shed further light on these data, we turned to SAPT,[247, 248] at the SAPT0/jun-cc-pVDZ level[230] using Psi4.[244] SAPT provides not only robust interaction energies but also decomposes these interactions into underlying physical components (electrostatic, exchange repulsion, induction, and dispersion interactions).[249] These interaction energy components are plotted as a function of the local orientation angle for stacked dimers of 1–4 in Figure 2c-2f.[250] Perhaps surprisingly, the incorporation of carbonyl groups into sumanene leads to a weakening of the electrostatic preference for staggered configurations. Whereas the electrostatic component favors the staggered configuration of 1 by ~7 kcal mol-1 compared to the eclipsed configuration, this difference is only 1 kcal mol-1 for 4. Also somewhat surprisingly, the exchange-repulsion component of these interactions strongly favors the eclipsed configuration for 1, seemingly at odds with proposed steric interactions between endo-H and exo-H atoms in the eclipsed configuration. These two trends arise primarily from variations in bowl to bowl distance (h) with the local orientation (see Figure 2b). For 1, the dimer is more compact in the staggered configuration than eclipsed (h = 3.81 Å for eclipsed vs. 3.72 Å for staggered), in order to avoid steric interactions between endo-H and exo-H atoms in the eclipsed conformation. This results in much stronger electrostatic, dispersion, and exchange-repulsion interactions in the staggered configuration and a total interaction energy that strongly favors the staggered configuration. In contrast, as a result of the lack of CH2 groups,[251] stacked dimers of 4 exhibit similar stacking distances in both eclipsed and staggered configurations (3.76 and 3.75 Å, respectively). This leads to the rather small difference in total interaction energies between the eclipsed and staggered

configurations.[252] With regard to the systematic increase in barrier height for conversion of eclipsed to staggered configurations with increased number of carbonyl groups, the SAPT data reveal that a decrease in the attractive part overwhelms the decrease of the repulsive interaction, leading to a gradual rise in the energy required to transform from eclipsed to staggered configurations.

Overall, analyses of pairwise stacking interactions are unable to fully explain the different orientations of 1-4 in the solid state, since, based on the above results and recent work from Chen et al.,[219] one would expect 1 – 4 to all pack in a staggered configuration in the solid state. Alternatively, three-body stacking terms could impact the preferred local orientation of 1-4. However, computed interaction energies for a model stacked trimer of 4 (see Appendix B Figure B-9) reveal that the staggered configuration is still the global minimum, lying 1.2 kcal mol-1 lower in energy than the eclipsed configuration. This is roughly double the energy difference in the stacked dimer of 4, indicating that many-body stacking interactions have no substantial effects. This suggests that the orientational preferences of extended stacks of these molecules will mimic those of the isolated stacked dimers.

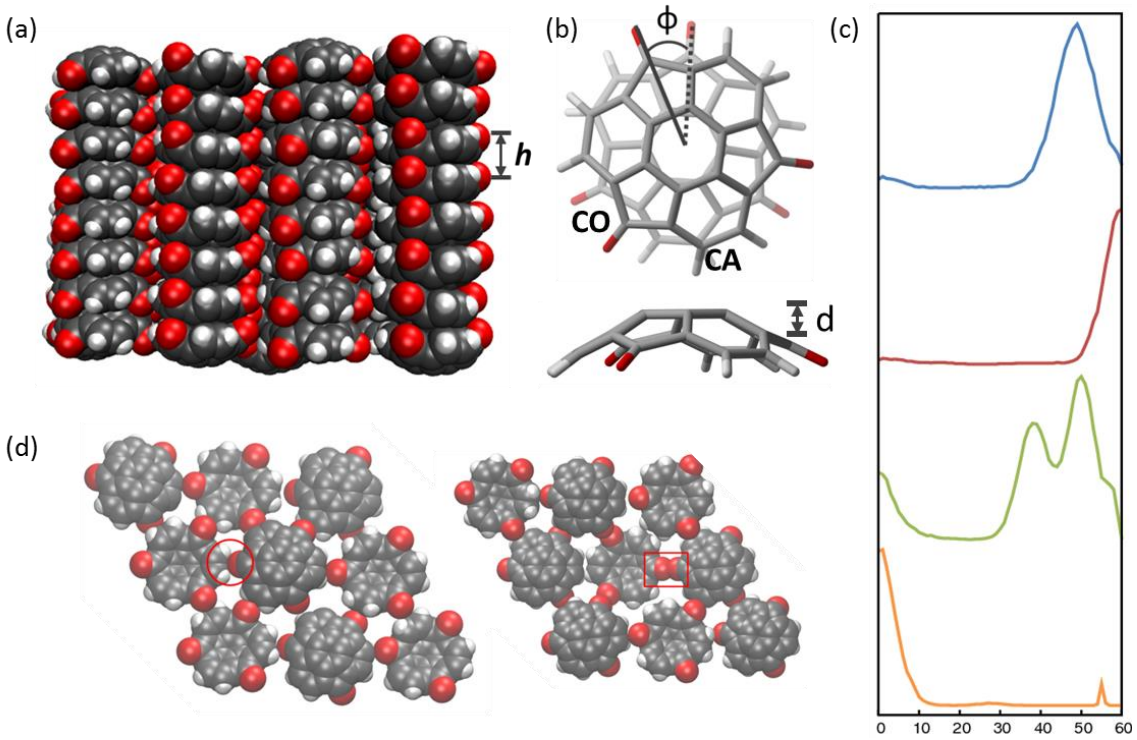


Figure IV-4. (a) Snapshot of 4 from the MD simulation; (b) definition of key intra- and intermolecular parameters; (c) distribution of local orientation angles averaged over 10ns MD simulations (from top to bottom, 1-4); (d) Unfavorable intercolumnar O···O interactions in the staggered configuration (right) and potentially favorable O···H interactions in the eclipsed configuration (left).

Instead, the observed eclipsed packing of 4 must arise from intercolumnar interactions. To evaluate these intercolumnar interactions, we turned to MD simulations, which have proven powerful in the study of columnar stacking of discotic molecules.[61, 210] However, to our knowledge there have been no previous MD simulations of sumanene derivatives in the solid state. MD simulations of systems comprising 128 molecules in 16 stacks of 8 molecules (see Figure 3) were done using GROMACS 4.6.5.[245] To justify our choice of MM potential and simulation protocol (see Appendix B for details), we compared intermolecular distances, bowl depths, and maxima in the distribution of local orientations in the solid state with experimental data for 1, 2, and 4

(see Table 1). Overall, structural parameters averaged over the 1 ns MD simulations are in good agreement with available data from X-ray analyses, suggesting that the MD simulation protocol used is a reliable tool to reveal details about the crystal packing of these systems. Figure 3c shows the distribution of local orientations over 1 ns of these simulations. Molecules 1 and 2 show strong tendencies to stack in staggered conformations, with maxima at ϕ = 50 and 58°, respectively. This is in good agreement with the X-ray derived angles of 55 and 60°.[214, 218] This can be compared to the data for 4, which strongly favors an eclipsed conformation (ϕ = 0°). This is again in agreement with experiment.[218] For 3, the predicted packing is more complicated in that three prominent peaks are present in the distribution, representing both staggered (resembling 1 and 2) and eclipsed (as in 4) orientations; the solid-state packing of 3 can be viewed as a transition from staggered conformation of 1 and 2 to the eclipsed one of 4.[253]

Table IV-1. Crystal structure data for 1-4 in the solid state along with the mean computed values from MD simulations.

	Calc.[a]				Expt[214, 218]			
	h(Å)	ϕ(°)	d_{CA}[b]	d_{CO}[c]	*h*	ϕ	d_{CA}	d_{CO}
1	3.86	50	1.21	0.91	3.86	55	1.11	0.90
2	3.77	58	1.11	-	3.82-3.87	60	1.11-1.13	-
3	3.82	~155	1.12	-	-	-	-	-
4	3.74	0	1.12	0.99	3.75	0	1.13	0.99

[a]Computed values are mean values averaged over 500 snapshots by a 1 ns data collecting run (see Appendix B).
[b]Bowl depth from the bottom of bowl to CH carbon CA, in Å.
[c]Bowl depth from the bottom of bowl to CO carbon CO, in Å.

These MD simulations reproduce the experiment finding of a transition from staggered to eclipsed packing upon incorporation of three carbonyl groups, supporting the importance of intercolumnar interactions in determining the packing of these bowl-shaped molecules. To further elucidate the origin of the different preferences for staggered and eclipsed stacking motifs, we again turned to DFT. This time, however, we incorporated additional, neighboring monomers. In particular, we combined the MD simulations with DFT computations by extracting information about the position and orientation of individual molecules in the solid from the MD simulation, while the geometry of molecules and interaction energies were computed using DFT. Eleven molecules with a central trimer and eight surrounding monomers from peripheral columns were extracted from the final snapshot of the MD simulations (see Appendix B Figure B-4). This cluster was selected to ensure that all monomers having non-negligible contacts with the central monomer were included. Then, each monomer was replaced with the DFT-optimized structure, keeping the center of mass and relative orientations of each monomer the same as the MD simulation. The interaction energy of the central molecule with the surrounding 10 molecules was computed at the B97D/TZV(2d,2p) level of theory for different relative orientations at 3° intervals.

The resulting interaction energies, which now include intercolumnar interactions, are drastically different from those of the isolated dimers (see Figure 2f), especially for 4. Overall, there is a tendency of increasing energy of the staggered configuration, relative to the eclipsed configurations, as seen for the isolated dimers (Figure 2a). Indeed, for 1, the staggered conformation is favored by 5.5 kcal mol-1 over the eclipsed

conformation, which is roughly twice the energy difference in the corresponding dimer. Since in this cluster model the central structure is a stacked trimer (see Appendix B Figure B-4), this suggests that intercolumnar interactions have little impact on the local orientations of sumanene 1 in the solid state. For 4, on the other hand, the eclipsed conformation is strongly favored, lying 25 kcal mol-1 lower in energy than the staggered configuration. This substantial energy gap highlights the importance of intercolumnar interactions in the molecular packing of 4. This can be explained by two complementary effects. First, as noted by Sakurai and co-workers,[218] there are potentially favorable CH…O interactions[254] in the eclipsed configuration. At the same time, there are unfavorable interactions between carbonyl groups in the staggered conformation (see Figure 3d). For 2 and 3, which lack C3-symmetry, consideration of angles up to 180° reveal global energy minima at 60° and 180° for 2 and 3, respectively (see Appendix B Figure B-13). Overall, these DFT data are consistent with experiment and the results of the MD simulations, suggesting that this 11-molecule cluster is sufficient to capture both the intra- and intercolumnar interactions responsible for the solid state packing of these bowl-shaped molecules.

4.4 Conclusions

Isolated stacked dimers and trimers of sumanene (1) and sumanene derivatives 2 – 4 prefer staggered configurations. It is not until intercolumnar interactions are included that the preference of sumanenetrione (4) to pack in an eclipsed configuration[218] is captured. SAPT computations revealed that the preference for isolated stacked dimers of

1 – 4 to adopt staggered orientations is driven by favorable electrostatic and dispersion interactions. In the case of sumanenetrione, this inherent tendency to adopt staggered configurations within a given column is overwhelmed by unfavorable O…O contacts between stacked columns in the solid state, leading to the observed transition from staggered to eclipsed packing motifs. Bowl-shaped molecules exhibit diverse packing motifs in the solid state, often confounding efforts to rationally design materials from such molecules with well-defined packing motifs.[208, 212, 213, 255] The present results constitute a key step toward understanding the packing of such molecules by demonstrating how inter- and intracolumnar interactions can be quantified through DFT and MM-based computational studies. Moreover, these results suggest that carefully designed intercolumnar interactions can be used to override the inherent tendencies of many discotic and bowl-shaped molecules to adopt staggered relative orientations within columnar stacks.

CHAPTER V

CONFORMATIONAL BEHAVIOR AND STACKING INTERACTIONS OF CONTORTED POLYCYCLIC AROMATICS*

5.1 Introduction

Organic electronic materials can be fabricated from various small molecules and often combine unique desirable properties including relative ease of processing, greater flexibility, and lower costs.[256] The electronic properties of these materials and their performance in devices depend strongly on the packing of the constituent molecules in the solid state. For example, discotic and curved molecules can assemble into columnar stacks arranged in a regular lattice that gives rise to useful properties for electronic applications, including the ability to conduct charges along the stacks.[257, 258] Even within such well-ordered packing motifs, subtle changes in relative orientations can have an enormous impact on material performance.[210] Understanding the various non-covalent interactions that govern the solid-state packing of curved and planar polycyclic aromatic molecules (*e.g*. stacking interactions, CH/π interactions, *etc*.) will aid in the rational design of materials with precisely tailored electronic properties. Non-covalent interactions between planar aromatic species have been extensively explored.[225, 259-264,265] However, our understanding of the molecular packing of curved polycyclic systems is less well-developed. The general concept of concave-convex stacking, in which

*Adapted with permission from "Conformational Behavior and Stacking Interactions of Contorted Polycyclic Aromatics", by Y. Guan, M. L. Jones, A. E. Miller, and S. E. Wheeler, 2017. *Phys. Chem. Chem. Phys.* **19**, 18186. Copyright 2017 Royal Society of Chemistry.

enhanced stacking interactions arise from the complementary polarization of the π-systems of the concave and convex faces of two curved systems, has long been discussed;[266, 267] however, there have been a limited number of computational studies of curved polycyclic species.[268, 269] Sherrill and co-workers[268] examined the impact of curvature on stacking interactions in model systems, reporting that nested stacking interactions are enhanced by increased curvature, except at short distances. This was explained by the increased molecular dipole moment that arises when polycyclic aromatic systems are curved.

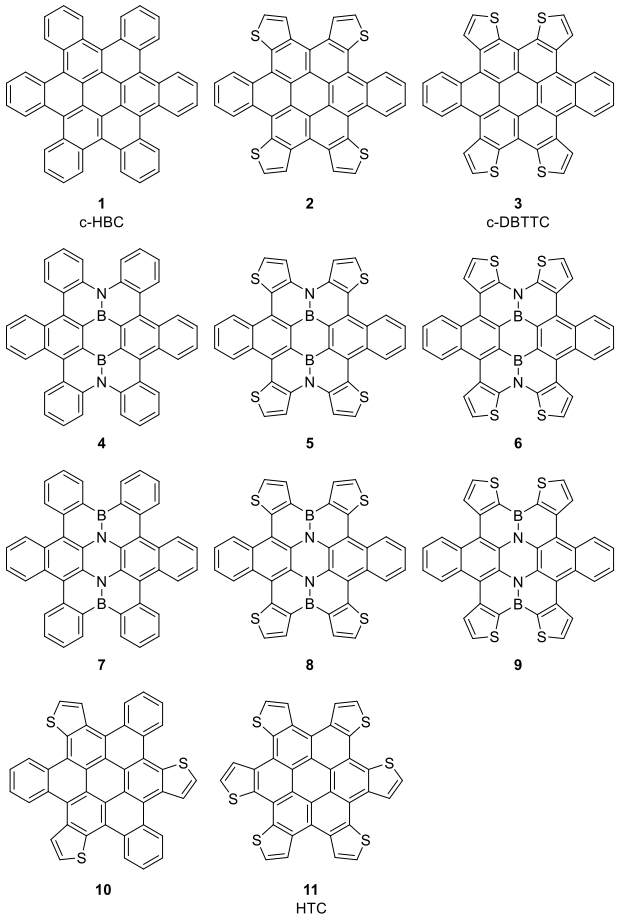
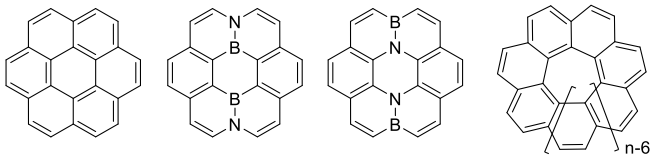


Figure V-1. Doubly-concave polycyclic aromatic systems related to c-HBC. Alkylated versions of 1, 3, 8, and 11 have been synthesized.[76, 270-274]



12 coronene ([6]circulene) **13** **14** **15-18** [n]circulene

Figure V-2. Coronene (12), its B-N substituted analogues (13-14), and the [n]circulenes (n = 7-10, 15-18).

Several major classes of curved molecules have attracted attention in the context of organic electronic materials, including bowl-shaped molecules,[208, 212, 255, 275] twisted molecules,[77, 276-280] and contorted (*i.e.* saddle-shaped or doubly-concave) molecules.[76, 270, 271, 281, 282] The unique shapes of contorted and bowl-shaped molecules, coupled with their ability to interact strongly with electron-acceptors such as fullerenes, portends their use in a variety of organic electronic materials.[283, 284] We recently studied the stacking interactions of bowl-shaped molecules (*i.e.* sumanene and its derivatives);[285] here, we consider stacking interactions of doubly-concave molecules, which offer an opportunity to tune the strength of stacking interactions by regulating the curvature of the molecular plane without the introduction of a net molecular dipole moment.[76, 270, 272]

Below, we consider a series of doubly-concave molecules related to the contorted hexabenzocoronene (c-HBC) structure of Nuckolls and co-workers,[270, 286] as well as the [n]circulenes[287-290] (See Figures 1 and 2). A number of substituted analogues of c-HBC have been synthesized, and many exhibit promising performance as organic electronic materials. For example, alkylated derivatives of **1**, **3**, **8**, and **11**,[76, 270-274] have been synthesized, demonstrating that contorted molecules can assemble into regular solid-state phases. For example, Nuckols *et al.*[291] reported three crystalline polymorphs in thin

films of **1** based on post-deposition processesing, while two crystal packing motifs were observed for **3** depending on the crystallization conditions.[76] Similarly, the crystal structure of alkylated versions of **8** was reported to exhibit a mixture of two conformations in an A-A-B-A-A-B order within columnar stacks.[274] Structures of the [n]circulenes have been studied computationally by a number of groups.[292-294] Deep saddle conformations were confirmed for monomers of [7]circulene (**15**) to [16]circulene.[295] However, stacked homodimers of these various saddle-shaped molecules have not previously been studied computationally.

A sound understanding of both the conformational behaviour and stacking interactions of these saddle-shaped molecules is vital for the rational design of contorted molecules for use in organic electronic materials. For many of these systems, there are multiple low-lying conformations, and the conformations and solid-state packing will impact the resulting optoelectronic properties. Here, we provide a systematic computational study of the conformational behaviour of these doubly-concave molecules as well as the π-stacking interactions of homodimers of these systems.

5.2 Computational Methods

Conformations of isolated molecules of **1**-**18**, as well as stacked homodimers of these molecules, were optimized using dispersion-corrected density functional theory at the B97-D/TZV(2*d*,2*p*) level of theory.[296-299] Conformations of **1**-**11** were identified by starting with a fully planar geometry of each molecule and then subsequently and repeatedly following all imaginary vibrational modes to lower-energy conformations.

All structures presented are energy minima, as confirmed by harmonic vibrational frequency analyses. Geometries of stacked dimers of **1-18** were explored by systematically considering all unique stacked homodimers of each low-lying conformation (see below for more details). Given the importance of dispersion interactions in these dimers, we explored several variants of Grimme's D3 empirical dispersion correction.[300] Based on its ability to reproduce the extrapolated CCSD(T) benchmark interaction energy for the coronene dimer from Janowski and Pulay,[301] we present data computed at the B97-D3M(BJ)/TZV(2*d*,2*p*) level of theory.[296-300, 302, 303] Binding energies, defined as the difference in electronic energy between the optimized stacked dimer and the lowest-energy conformation of the isolated monomers, are reported for stacked dimers lying within 3 kcal mol^{-1} of the global minimum energy structure. All computations used Gaussian09[77] and employed density fitting techniques.

5.3 Results and Discussion

5.3.1. Conformations of contorted polycyclic aromatics

Although previous experimental and computational work has unveiled information regarding the possible conformations of some of the systems depicted in Fig. 1,[76, 270, 272, 274] the conformational behaviour of many of these systems has not previously been explored. To provide a more comprehensive understanding of isolated molecules of **1-11**, we studied conformations within 3 kcal mol^{-1} of the global minimum energy conformation. Molecules **1-11** exhibit four distinct types of low-lying conformations (see Figure 3). For molecules with more than three peripheral benzo rings

(*i.e.* **1**, **4**, **7**, and **10**) only the up-down-up-down (UD) conformation is thermodynamically accessible due to the highly crowded peripheral environment. On the other hand, when some of these rings are replaced by thiophenes, additional low-lying conformations become accessible, including saddle (S), twisted saddle (TS), and twist-down-up-twist-down-up (TD) conformations. For most of these systems, the UD conformation is preferred. However, for **3**, **8**, and **9** the S conformation is lower-lying, with the UD conformation still thermodynamically accessible. For **11**, the TS and UD conformations are similar in energy, with the TS conformation lying 0.3 kcal mol^{-1} lower than S. The low-lying conformations of these molecules vary in their degree of deviation from planarity. Table 1 lists two angles that quantify the extent of contortion (φ and θ; see Figure 3), along with the corresponding energies relative to the lowest-lying conformation of each species. Energies of the corresponding planar conformations, relative to the lowest-lying conformation, are also listed to quantify the overall strain relief achieved by adopting contorted conformations.

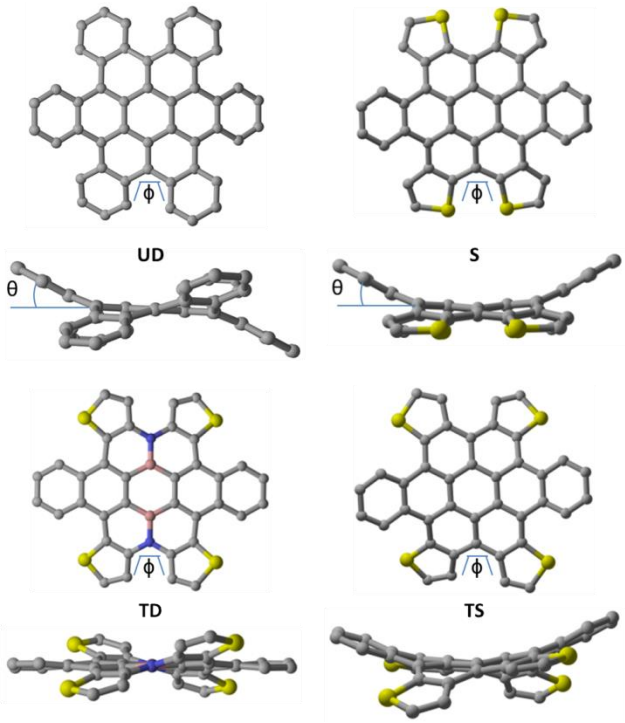


Figure V-3. Unique optimized conformations of 1-11, Up-Down-Up-Down (UD), Saddle (S), Twist-Down-Up-Twist-Down-Up (TD), and Twisted Saddle (TS). Two angles describing the degree of contortion are also presented: φ, the angle between two adjacent overlapping anthradihiophenes or pentacenes and θ, the tilt angle of the benzo ring with respect to central coronene core. Hydrogen atoms have been removed for clarity.

A number of trends emerge from the data in Table 1. First, as noted by Nuckolls and co-workers,[76] replacing benzo rings in c-HBC with thiophenes (*i.e.* **2** and **3**) results in considerably more planar and more flexible structures. For example, for **1**, the bending angle of pentacene (θ) is 23°. For **2** and **10**, this angle is reduced to 16° and 20°, respectively. When all six benzo rings are replaced by thiophenes (*i.e.* **11**), the low-lying conformation is very nearly planar, with the fully-planar conformation lying only 3.8 kcal mol^{-1} higher in energy. For tetrathieno-substituted molecules **2** and **3**, the orientation of the thiophenes impacts the structures considerably. In particular, the

preferred conformation of **2** is UD, whereas the S conformation is lowest-lying for **3**. This can be explained in terms of the steric hindrance between the hydrogens of neighboring thiophene groups. The result is that the S conformation of **2** is twisted and higher in energy than the UD conformation by 2.2 kcal mol^{-1}. However, for **3** these steric interactions are absent and the saddle conformation becomes the lowest-lying conformation.

Table V-1. Low-lying conformations of 1-11 as well as a constrained planar conformation. Energies (E_{rel}, in kcal mol^{-1}) are reported relative to conformer 1.[a]

	Conformer 1			Conformer 2				Planar
Mol	φ	θ	Conf	φ	θ	E_{rel}	Conf	E_{rel}
1	43	23	UD	-	-	-	-	154.4
2	33	16	UD	22	18	2.2	TS	21.1
3	0	22	S	25	20	1.5	UD	31.3
4	48	22	UD	-	-	-	-	140.8
5	37	14	UD	38	0[b]	1.3	TD	20.2
6	30	18	UD	0	20	0.7	S	23.5
7	41	23	UD	-	-	-	-	153.5
8	0	21	S	32	19	1.7	UD	25.7
9	0	22	S	28	21	1.4	UD	41.2
10	34	20	UD	-	-	-	-	43.0
11	5	15	TS	28	11	0.3	UD	3.8

[a] See Figure 3 for definitions of the angles φ and θ (given in degrees).
[b] Twisted angle of pentacene is 11° (see Appendix C for details).

Computations also reveal subtle differences in the structures of these systems depending on whether the N or B is located closer to the centre of the coronene (*e.g.* **4** vs **7**). In particular, in **4-6**, there is greater deviation from planarity with regard to the thiophene part (larger φ) and more planar pantacene components (smaller θ), compared to the corresponding structures among **7-9**.

5.3.2. Stacking of coronene and its BN-analogues

Molecules **1**-**11** contain either a coronene (**12**) or B-N substituted coronene core (**13**-**14**). We first considered the stacking interactions of these planar systems to provide a baseline for understanding their benzanulated and thiophenylated derivatives (**1**-**11**). Low-lying stacked dimers of **12**-**14**, along with computed binding energies, are displayed in Figure 4. Intermolecular coordinates including stacking distance (h), displacement distance (Δd), and orientation angle (α) are also provided. Stacked dimers of these planar systems can be in either a parallel-displaced configuration, in which the centroid of one molecule is displaced parallel to the molecular plane of the other molecule (Δd >> 0), or a sandwich configuration in which the rings share a common symmetry axis (Δd = 0). The lowest-lying configuration for coronene (**12**) is the parallel-displaced configuration, which is lower than the sandwich configuration by 1.1 kcal mol^{-1}. For **13** and **14**, in which B and N atoms have been embedded in coronene, the lowest-lying stacked dimer is in a sandwich configuration. The most favourable parallel-displaced configurations are 0.5 and 1.0 kcal mol^{-1} higher in energy for **13** and **14**, respectively. In the low-lying sandwich dimers, one monomer is rotated by either 30 or 90° with respect to the shared symmetry axis. In both **13** and **14**, orientation of 30° is slightly lower in energy than that at 90°. Such sandwich stacking configurations are reminiscent of those seen for many discotic systems.[210, 304]

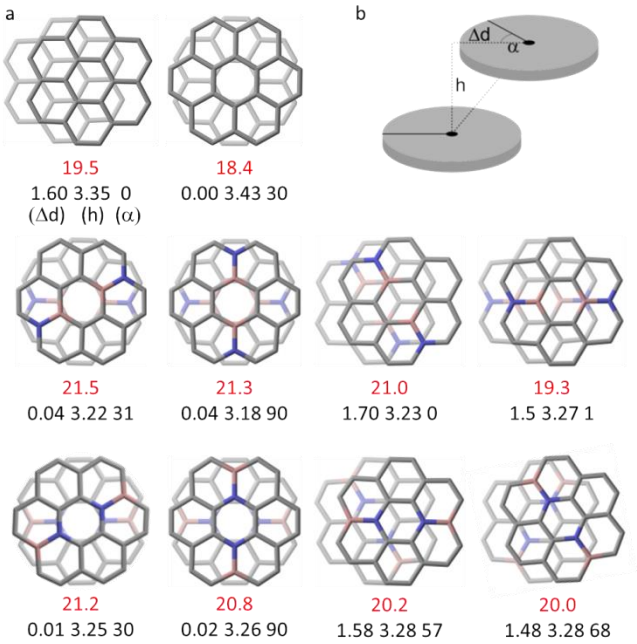


Figure V-4. (a) Low-lying stacked dimers of 12-14 and associated binding energies (red, in kcal mol^{-1}) and intermolecular coordinates. Hydrogen atoms have been removed for clarity. (b) Definition of intermolecular stacking coordinates: stacking distance (h), displacement distance (Δd), and orientation angle (α).

5.3.3. Stacking of contorted polycyclic aromatics

To understand the stacking interactions of contorted molecules, we first extended coronene (**12**) to a hypothetical planar version of c-HBC (see Figure 5a). The interaction energy of a stacked dimer of this hypothetical planar c-HBC was then computed as a function of h and Δd (see Figure B-3). A parallel-displaced configuration was identified as the minimum energy structure (see Figure 5a) with h = 3.25 Å and Δd = 1.6 Å. The corresponding interaction energy, relative to the isolated planar monomers, is 42.9 kcal mol^{-1}. This highly favourable interaction energy, which is more than twice that of the coronene dimer, can be attributed to the introduction of favourable parallel displaced stacking geometries[305] between the peripheral benzo rings.

This idealized stacking interaction between hypothetical planar versions of c-HBC can be contrasted with the lowest-lying stacked conformation of the relaxed (i.e. contorted) c-HBC dimer (see Figure 5b). Overall, the interaction energy is 8 kcal mol^{-1} less favourable for the stacked dimer of contorted c-HBC compared to the hypothetical planar version. This reduced interaction energy runs counter to the concept that increasing curvature in nested π–π interactions leads to more favourable stacking energies,[268] and can be attributed to a number of competing effects. First, for the contorted c-HBC dimer the coronene cores adopt an eclipsed sandwich configuration, which, at this stacking distance, results in a 3.6 kcal mol^{-1} loss in interaction energy compared to the more favourable parallel-displaced configuration (see Figure B-4). Two other, inter-related factors also reduce the interaction energy: intermolecular steric hindrance between the peripheral benzo rings and a larger stacking distance between the coronene cores in the case of the contorted c-HBC. These effects are partially offset by favourable intermolecular interactions of the peripheral benzo rings, which exhibit nearly ideal parallel displaced stacking interactions.[305] Finally, comparing the binding energy for the c-HBC dimer (33.5 kcal mol^{-1}) with the interaction energy, we see that there is 1.4 kcal mol^{-1} of strain energy in the two c-HBCs in the stacked dimer. This results from distortions of the two monomers, as reflected in the difference in distances between the coronene cores (3.62 Å) and between the centroids of the benzo rings (4.04 Å).

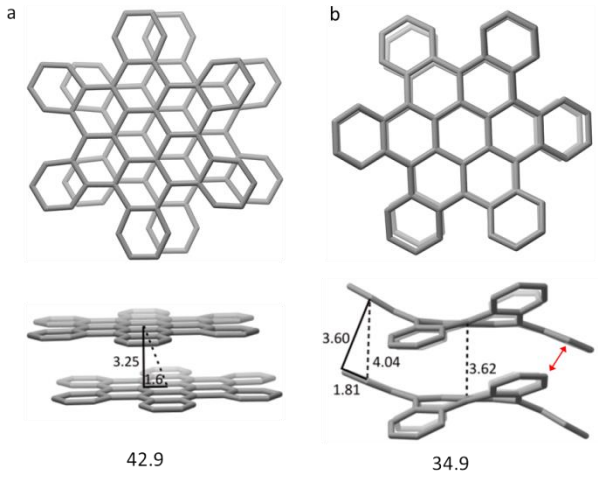


Figure V-5. (a) Stacked dimer of a hypothetical planar version of c-HBC, along with the interaction energy in kcal mol^{-1}. (b) Fully optimized stacked dimer of c-HBC (UD1 configuration; see Figure 6), along with the interaction energy in kcal mol^{-1}. The red arrow indicates one of six intermolecular steric interactions between benzo rings. Hydrogen atoms have been removed for clarity.

The two stacked dimers shown in Figure 5 can be considered limiting cases for stacking interactions of the contorted molecules **2-11**, since the curvature of these other systems lie between those two extremes. That is, as the monomers become more planar, we anticipate a gradual shift from the eclipsed stacking seen in Figure 5b to a parallel displaced configuration as in Figure 5a. However, the stacking interactions of **2-11** are slightly more complex, given the presence of heteroatoms, the existence of different monomer conformations, and the possibility of multiple distinct stacking configurations depending on the symmetry of the monomers.

The possible stacking modes that can be adopted by homodimers of **1-11** are represented in Figure 6 for molecules in up-down-up-down (**UD1**-**UD6**) and saddle (**S1**-**S2**) conformations. Depending on the symmetry of a given conformation, one or more

distinct stacked dimers can be built. For example, the UD conformation of c-HBC is of D_{3d} symmetry, for which only one stacking configuration (**UD1**) can be constructed (see Figure 5b). In contrast, for the chiral UD conformation of **10**, which has C_3-symmetry, four unique configurations (**UD1**, **UD3**, **UD4**, and **UD5**) are possible. Data for the most energetically favourable stacked dimers of **1**-**11** are listed in Table 2. In particular, we present binding energies for dimers of each of the low-lying conformers, relative to the lowest-lying conformation of the separated monomers, as well as relative energies for the different configurations of a given contorted molecule. Intermolecular coordinates characterizing these complexes are also listed, as defined in Figure 4b.

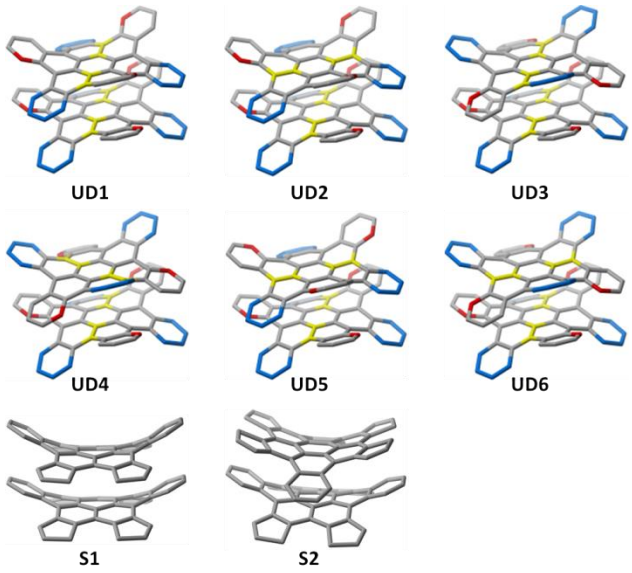


Figure V-6. Classification of different configurations for stacked dimers of contorted molecules in Up-Down-Up-Down (UD1-UD6) and Saddle (S1-S2) conformations. Hydrogen atoms have been removed for clarity.

Representative structures of favourable stacked dimers are presented in Figure 7 (coordinates for all computed structures can be found in SI). From the data in Tables 1 and 2, we see that the lowest-lying monomer conformation always leads to the most energetically favourable dimer. That is, the most favourable stacked dimers of **3**, **8**, **9**, and **11** feature saddle-like conformations, whereas the other lowest-lying stacked dimers are for UD conformations. Compared to dimers of **1**, the introduction of either thiophene rings or embedded N-B atoms lead to enhanced stacking interactions in all but one case (**7**). These dimers exhibit a mixture of sandwich ($\Delta d \approx 0$) and parallel-displaced configurations ($\Delta d >> 0$), as well as both eclipsed ($\alpha \approx 0$) and staggered ($\alpha \approx 60$) orientations. However, within these data there are a number of general trends.

First, considering only the UD1 conformations of these stacked dimers, there is still a mixture of sandwich and parallel-displaced configurations (by construction, UD1 will always be eclipsed). In general, the preference for sandwich configurations is associated with greater curvature of the monomers, larger stacking distances (h), and weaker overall binding (*e.g.* UD1 dimers of **1, 2**, **7**, and **10**). For other UD1 dimers, the stacking is more compact and parallel-displaced configurations are preferred. For instance, for **5** and **11,** which are considerably more planar than **1**, the stacking distances (3.23 and 3.34 Å, respectively) are similar to those seen for the hypothetical planar c-HBC dimer (Figure 5a). These two dimers exhibit large parallel displacements (Δd = 1.24 Å in both cases) and binding energies far exceeding 40 kcal mol^{-1}.

Table V-2: Binding energies (relative to the most stable conformation of the corresponding monomers) as well intermolecular coordinates of low-lying conformations of stacked dimers of 1-11.[a]

Mol	Config.	E_{bind}	E_{rel}	Δd	h	α
1	UD1	33.5	0.0	0.00	3.62	0
2	UD1	40.6	0.0	0.12	3.53	0
	S2	40.5	0.1	0.25	3.40	70
	S1	39.5	1.1	0.27	3.51	8
3	S1	40.5	0.0	0.86	3.55	5
	UD1	38.7	1.8	0.93	3.42	0
	S2	38.5	2.0	0.26	3.52	47
	UD2	38.1	2.4	0.06	3.53	59
4	UD2	36.5	0.0	0.45	3.28	60
	UD1	36.3	0.2	0.78	3.32	0
5	UD1	44.3	0.0	1.24	3.23	2
	UD2	43.5	0.8	0.67	3.27	58
6	UD2	42.9	0.0	0.56	3.28	61
	UD1	40.5	2.4	0.90	3.30	1
7	UD2	32.5	0.0	0.10	3.60	60
	UD1	32.2	0.3	0.06	3.58	1
8	S2	44.2	0.0	0.36	3.39	74
	S1	42.0	1.8	0.34	3.48	12
9	S2	39.7	0.0	0.32	3.48	75
	S1	38.6	1.1	0.34	3.52	1
	UD2	38.0	1.7	0.03	3.49	59
	UD1	37.4	2.3	0.88	3.40	0
10	UD5	42.0	0.0	0.02	3.59	60
	UD4	40.6	1.4	0.00	3.53	56
	UD3	40.5	1.5	0.00	3.60	60
	UD1	39.5	2.5	0.00	3.54	1
11	S1	43.6	0.0	0.23	3.49	4
	UD1	43.4	0.2	1.24	3.34	0
	UD3	42.2	1.4	1.21	3.32	0
	UD6	41.4	2.2	1.23	3.33	60
	UD2	41.0	2.6	1.16	3.36	60

[a] Stacked dimers within 3.0 kcal mol^{-1} of the global minimum energy structure are shown.

For the saddle-like dimers (**S1** and S2), one important trend that emerges is that thiophene-containing c-HBC derivatives with sulfurs facing away from each other engage in stronger stacking interactions than those in which the sulfurs face toward each other (*e.g.* **2** vs **3** and **8** vs **9**).[306] Close examination of these saddle-like dimers reveals that the highly favourable stacking interactions exhibited by **2** and **8** may derive from the more favourable interactions between peripheral thiophenes (see Figure 7b). For example, comparing **8** and **9**, the symmetry and structures of these molecules (which is impacted by the different steric interactions around the periphery of each monomer) lead to a much closer, head-to-head stacking interaction between thiophenes in the former case, compared to the more distant tail-to-tail interaction in the latter case. The former interaction between thiophenes has been demonstrated to be the low-lying displaced stacking mode for thiophene dimers.[307]

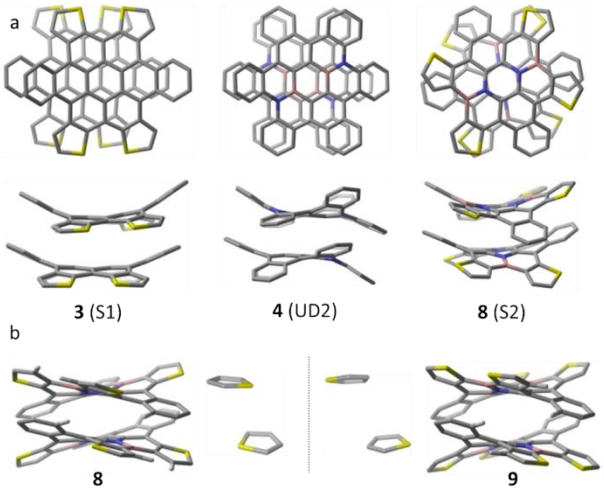


Figure V-7. a) Representative lowest-lying conformations of stacking dimers of 1-11; b) comparison of stacked dimers of 8 and 9. Also shown are the relative orientations of the stacked thiophenes present in these two dimers. Most hydrogen atoms have been removed for clarity.

Overall, most of these contorted molecules exhibit two or more thermodynamically accessible stacking motifs, which, when combined with the myriad intermolecular interactions occurring in the solid state, will likely lead to complex packing arrangements. For instance, for **10** and **11**, which contain three and six thiophene rings, respectively, there are four unique UD dimers that can be constructed and all four are thermodynamically accessible. This presumably underlies the observation of Mullen and co-workers[272] of different disordered polymorphs of **11**. The existence of multiple polymorphs of these contorted molecules in the solid phase is more universal than can be seen solely from considering isolated stacked dimers. For instance, even for **1** (c-HBC), for which we predict a single thermodynamically accessible dimer configuration, multiple crystalline polymorphs can be observed,[291] including many with non-stacked configurations. Similarly, while isolated dimers of **8** are predicted to exhibit a strong preference for saddle stacking configurations, as noted above the crystal structure of alkylated versions of **8** exhibit a mixture of conformations within columnar stacks.[274] While such packing motifs cannot be explained by the present results, we note that the A-A portions of this packing motif feature **8** in the S2 configuration identified here as being the most favourable.

5.3.4. Stacking of [n]circulenes

Finally, we turn to the [n]circulenes, which provide a somewhat simpler framework for probing the impact of inherent curvature on stacking interactions.

Although the circulenes all adopt a single saddle-shaped conformation, a number

of potential non-covalent dimer configurations are possible; the relative stability of these configurations varies with the size and curvature of the systems. Optimized dimers of **15**-**18** are shown in Figure 8, along with predicted binding energies. For **15**, two stacked energy minima were identified corresponding to parallel displacements along different directions, with the former being more favourable by 2.2 kcal mol^{-1}.

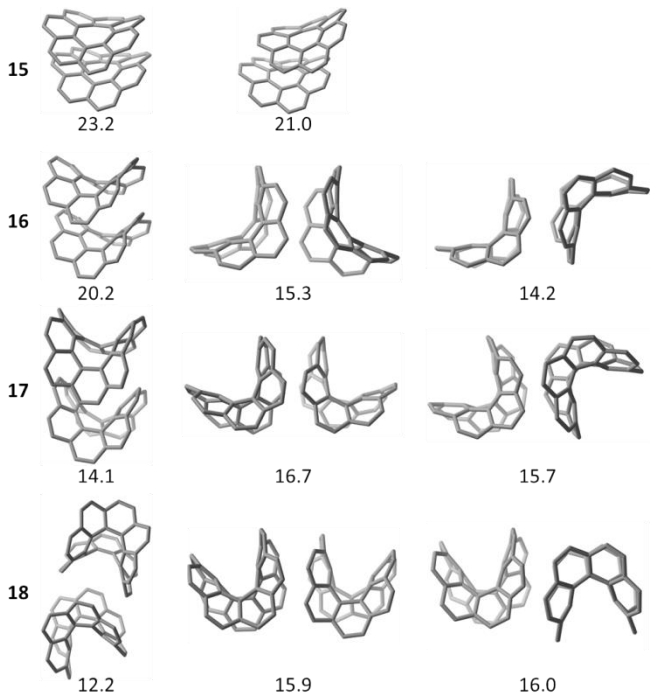


Figure V-8. Non-covalent dimers of [n]circulenes 15-18 along with computed binding energies in kcal mol^{-1}. Hydrogen atoms have been removed for clarity.

For the larger [n]circulenes (n > 7), the dimers fall in three main classes: the nested stacked configuration (*e.g.* the lowest-lying dimer of **16**), the edge-to-edge (or elephant seal)[308] configuration (*e.g.* the lowest-lying dimer of **17**), and the up-down edge-to-edge configuration (*e.g.* the lowest-lying dimer of **18**). Considering only the nested dimer configuration, the curvature exhibited by the [n]circulenes initially leads to

enhanced stacking interactions (*i.e.* the binding energy is more favourable for **15** and **16** than for coronene). However, with further increased curvature there is a rapid decline in the strength of stacking interactions. For **18**, the curvature is so severe that close association is not possible in this nested stacking configuration and the corresponding binding energy is only 12.2 kcal mol^{-1}. Second, turning to the two edge-to-edge configurations, both elephant seal and up-down edge-to-edge configurations become possible with **16**. As the size of the circulenes increases, the binding energy in the elephant seal configuration is maximized at **17**, while that of the up-down edge-to-edge configuration continues to increase through **18**. The net result is a gradual transition of the lowest-lying configuration from nested to elephant seal to up-down edge-to-edge going from **15** to **18**. This trend is a simple consequence of the increasing curvature of the successively larger circulenes. The structures of [9]circulene (**17**) and [10]circulene (**18**) are too contorted to provide sufficient contact surface area in the nested configuration; however, the sides of **17** and **18** are flatter then the top surface, allowing for more significant surface area contact in the two edge-to-edge configurations. Moreover, the two edge-to-edge configurations exhibit favourable intermolecular T-shaped interactions[225, 309, 310] that further enhance their interaction.

Considering the lowest-lying configuration for each system, the overall trend is a gradual decrease in binding energy going from **15** to **18**. In other words, whereas curvature typically enhances stacking interactions, overly curved molecules do not always exhibit sufficiently complementary shapes to allow for strong binding.

5.4 Conclusions

Stacking interactions between curved polycyclic aromatic hydrocarbons represent a key frontier in the march toward a general understanding of stacking interactions in aromatic systems. Among curved systems, doubly-concave or saddle-shaped molecules present a particularly attractive area of study, since they can introduce curvature into stacking interactions in the absence of a net molecular dipole moment. We presented dispersion-corrected DFT data on the low-lying conformations of 18 saddle-shaped molecules as well as their stacked dimers. In general, the introduction of thiophene units and B/N substitutions into c-HBC results in stronger predicted stacking energies, and nearly all of the contorted polycyclic systems considered exhibit multiple thermodynamically accessible dimers. Among the [n]circulenes, the preferred interaction mode changes with the size of the system. [7]circulene stacks most favourably, and is the only curved circulene to stack more strongly than the planar coronene (*i.e.* [6]circulene); the excessive curvature of the larger circulenes precludes close association in a nested configuration, and the larger circulenes favour other stacking geometries.

Together, these data provide insights into the propensities of these saddle-shaped molecules to form stacked dimers, and serve as a reminder that the impact of curvature on stacking interactions is complex. Most importantly, the introduction of curvature does not always lead to stronger binding. Ultimately, even though the solid state packing of these contorted molecules cannot be explained based solely on isolated stacked homodimers, the present results constitute a key step toward a better understanding of the supramolecular assembly and packing of these molecules.

CHAPTER VI

SUMMARY AND CONCLUDING REMARKS

This dissertation highlights the significant role of computational chemistry in understanding and predicting properties and functions of molecular systems, especially asymmetric catalysis and polycyclic aromatic hydrocarbon self-assembly systems. Overall, our computed results are in very good agreement with the experimental data, upon which our computational model can further elucidate the mechanism of the system or propose novel structures. For the asymmetric catalysis projects, our computational toolkit opens the door to the automated computational catalysts design, while in the curved polycyclic aromatic hydrocarbon projects, our study identifies the key interactions governing packing organization as well as the complex relationship between stacking interactions and molecular structures.

In the first part of this dissertation, we presented our work on developing a computational toolkit that automatically screens potential catalysts for a given asymmetric reaction as well as applications of that toolkit on transition metal or pure organo- catalysis. First, we explained the detailed workflow of our computational toolkit, AARON, that can 1) automatically build TS guesses structures involving different configurations for catalytic reactions, 2) locate TS structures with precise error control and geometry vetting, 3) search conformers in a parallel and hierarchical way, and 4) predict selectivities through a Boltzmann weighting of all accessible TSs or reaction paths. Example applications demonstrated that AARON is effective for

predicting selectivities for complex reactions involving multiple key TS structures or stereodetermining steps.

Next, we have shown a detailed example of using AARON to study mechanisms of Rh-catalyzed asymmetric hydrogenation reactions for five C_2-symmetric phosphorus ligands. For each ligand there are two key steps, each containing mutliple configurations and conformations. To this end, AARON optimized 251 TSs across those two steps. The predicted selectivities agreed very well with the experimental data. Free energy profiles from these reactions showed that the mechanism varies with the different ligands, in contrast with decades of work on these reactions. Inspired by the critical non-covalent interaction operative in the key transition states, we designed a new ligand for this reaction predicted to provide both higher selectivity and reactivity.

In the second part of this dissertation, we turned to packing of curved polycyclic aromatic hydrocarbons. First, Density functional theory, symmetry-adapted perturbation theory, and classical molecular dynamics simulations were used to understand the local orientations within columnar stacks of four bowl-shaped molecues, sumanene, sumaneneone, sumanenedione, and sumanenetrione. Reliable quantum chemical computations on stacked dimers and trimers of these molecules reveal that all four systems prefer staggered configurations. The tendency of sumanenetrione to pack in an eclipsed configuration in the bulk materials was explained based on repulsive intercolumnar O···O interactions that override the inherent tendency of stacked columns to be staggered. Sumanenedione, for which crystal structure data are not available, was predicted to exhibit columnar packing with mixed staggered and eclipsed orientations.

Next, we present a systematic computational analysis of the conformations and stacking interactions of a set of 18 saddle-shaped, contorted polycyclic aromatic compounds. Computations reveal wide variations in both the nature of the low-lying conformations and the stacking affinities of these systems. In particular, the introduction of both thiophene rings around the periphery of these systems and the incorporation of B and N atoms into the coronene core can greatly enhance their tendency to form strongly stacked dimers. Overall, these data provide a reminder that curvature does not always lead to stronger stacking interactions.

REFERENCES

1. Antony, J.; Grimme, S., *Physical Chemistry Chemical Physics* **2006,** *8* (45), 5287-5293.

2. Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. R., *Journal of computational chemistry* **2007,** *28* (2), 555-569.

3. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., *The Journal of chemical physics* **2010,** *132* (15), 154104.

4. Goerigk, L.; Grimme, S., *Physical Chemistry Chemical Physics* **2011,** *13* (14), 6670-6688.

5. Moellmann, J.; Grimme, S., *The Journal of Physical Chemistry C* **2014,** *118* (14), 7615-7621.

6. Sure, R.; Antony, J.; Grimme, S., *The Journal of Physical Chemistry B* **2014,** *118* (12), 3431-3440.

7. Grimme, S.; Hujo, W.; Kirchner, B., *Physical Chemistry Chemical Physics* **2012,** *14* (14), 4875-4883.

8. Hujo, W.; Grimme, S., *Physical Chemistry Chemical Physics* **2011,** *13* (31), 13942-13950.

9. Halpern, J.; Trost, B. M., *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (15), 5347-5347.

10. Knowles, W. S., *Angewandte Chemie International Edition* **2002,** *41* (12), 1998-2007.

11. Noyori, R., *Angewandte Chemie International Edition* **2002,** *41* (12), 2008-2022.

12. Hayashi, T.; Kumada, M., *Accounts of Chemical Research* **1982,** *15* (12), 395-401.

13. Vineyard, B. D.; Knowles, W. S.; Sabacky, M. J.; Bachman, G. L.; Weinkauff, D. J., *Journal of the American Chemical Society* **1977,** *99* (18), 5946-5952.

14. Luo, S.-X.; Engle, K. M.; Dong, X.; Hejl, A.; Takase, M. K.; Henling, L. M.; Liu, P.; Houk, K. N.; Grubbs, R. H., *ACS Catalysis* **2018**, 4600-4611.

15. Liu, M.; Yang, P.; Karunananda, M. K.; Wang, Y.; Liu, P.; Engle, K. M., *Journal of the American Chemical Society* **2018**.

16. Lu, G.; Liu, R. Y.; Yang, Y.; Fang, C.; Lambrecht, D. S.; Buchwald, S. L.; Liu, P., *Journal of the American Chemical Society* **2017,** *139* (46), 16548-16555.

17. Omer, H. M.; Liu, P., *Journal of the American Chemical Society* **2017,** *139* (29), 9909-9920.

18. He, G.; Lu, G.; Guo, Z.; Liu, P.; Chen, G., *Nature Chemistry* **2016,** *8* (12), 1131.

19. Huang, G.; Liu, P., *ACS Catalysis* **2016,** *6* (2), 809-820.

20. Zhou, Q. L., *Angewandte Chemie International Edition* **2016,** *55* (18), 5352-5353.

21. Yu, X.; Wang, W., *Organic & biomolecular chemistry* **2008,** *6* (12), 2037-2046.

22. Enders, D.; Niemeier, O.; Henseler, A., *Chemical Reviews* **2007,** *107* (12), 5606-5655.

23. Ryan, S. J.; Candish, L.; Lupton, D. W., *Chemical Society Reviews* **2013,** *42* (12), 4906-4917.

24. Fu, Z.; Xu, J.; Zhu, T.; Leong, W. W. Y.; Chi, Y. R., *Nature Chemistry* **2013,** *5* (10), 835-839.

25. List, B., *Chemical Reviews* **2007,** *107* (12), 5413-5415.

26. Parmar, D.; Sugiono, E.; Raja, S.; Rueping, M., *Chemical Reviews* **2014,** *114* (18), 9047-9153.

27. Maji, R.; Mallojjala, S. C.; Wheeler, S. E., *Chem. Soc. Rev.* **2018,** *47* (4), 1142-1158.

28. Nimmagadda, S. K.; Mallojjala, S. C.; Woztas, L.; Wheeler, S. E.; Antilla, J. C., *Angewandte Chemie International Edition* **2017,** *56* (9), 2454-2458.

29. Atodiresei, I.; Schiffers, I.; Bolm, C., *Chemical Reviews* **2007,** *107* (12), 5683-5712.

30. Lam, Y.-h.; Grayson, M. N.; Holland, M. C.; Simon, A.; Houk, K. N., *Accounts of Chemical Research* **2016,** *49* (4), 750-762.

31. Sunoj, R. B., *Accounts of Chemical Research* **2016,** *49* (5), 1019-1028.

32. Peng, Q.; Paton, R. S., *Accounts of Chemical Research* **2016,** *49* (5), 1042-1051.

33. Wheeler, S. E.; Seguin, T. J.; Guan, Y.; Doney, A. C., *Acc. Chem. Res.* **2016,** *49* (5), 1061-1069.

34. Curtin, D. Y., *Rec. Chem. Prog.* **1954,** *15*, 111−128.

35. Poree, C.; Schoenebeck, F., *Accounts of Chemical Research* **2017,** *50* (3), 605-608.

36. Guan, Y.; Wheeler, S. E., *Angewandte Chemie* **2017**.

37. Lu, T.; Porterfield, M. A.; Wheeler, S. E., *Organic Letters* **2012,** *14* (20), 5310-5313.

38. Hansen, E.; Rosales, A. R.; Tutkowski, B.; Norrby, P.-O.; Wiest, O., *Accounts of Chemical Research* **2016,** *49* (5), 996-1005.

39. Preshlock, S. M.; Ghaffari, B.; Maligres, P. E.; Krska, S. W.; Maleczka, R. E.; Smith, M. R., *Journal of the American Chemical Society* **2013,** *135* (20), 7572-7582.

40. Friedfeld, M. R.; Shevlin, M.; Hoyt, J. M.; Krska, S. W.; Tudge, M. T.; Chirik, P. J., *Science* **2013,** *342* (6162), 1076-1080.

41. Santanilla, A. B.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P., *Science* **2015,** *347* (6217), 49-53.

42. Eksterowicz, J. E.; Houk, K., *Chemical Reviews* **1993,** *93* (7), 2439-2461.

43. Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O., *Journal of Chemical Theory and Computation* **2008,** *4* (8), 1313-1323.

44. 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., *Journal of chemical theory and computation* **2017,** *13* (11), 5780-5797.

45. Zimmerman, P. M., *The Journal of chemical physics* **2013,** *138* (18), 184102.

46. Vaucher, A. C.; Reiher, M., *Journal of chemical theory and computation* **2018**.

47. The code is available on GitHub: https://github.com/ComputationalCatalystDesign/Aaron

48. Wu, J.; Pisula, W.; Müllen, K., *Chemical Reviews* **2007,** *107* (3), 718-747.

49. Amaya, T.; Hirao, T., *Chemical Communications* **2011,** *47* (38), 10524-10535.

50. Rieger, R.; Müllen, K., *Journal of Physical Organic Chemistry* **2010,** *23* (4), 315-325.

51. Wang, X.; Zhi, L.; Tsao, N.; Tomović, Ž.; Li, J.; Müllen, K., *Angewandte Chemie International Edition* **2008,** *47* (16), 2990-2992.

52. Wang, J.; Wu, H.; Jin, L.; Zhang, J.; Yuan, Y.; Wang, P., *ChemSusChem* **2017**.

53. Roberson, L. B.; Kowalik, J.; Tolbert, L. M.; Kloc, C.; Zeis, R.; Chi, X.; Fleming, R.; Wilkins, C., *Journal of the American Chemical Society* **2005,** *127* (9), 3069-3075.

54. Mori, T.; Takeuchi, H.; Fujikawa, H., Field-effect transistors based on a polycyclic aromatic hydrocarbon core as a two-dimensional conductor. AIP: 2005.

55. Bayn, A.; Feng, X.; Müllen, K.; Haick, H., *ACS applied materials & interfaces* **2013,** *5* (8), 3431-3440.

56. Hashimoto, S.; Ikuta, T.; Shiren, K.; Nakatsuka, S.; Ni, J.; Nakamura, M.; Hatakeyama, T., *Chemistry of Materials* **2014,** *26* (21), 6265-6271.

57. Feng, X.; Marcon, V.; Pisula, W.; Hansen, M. R.; Kirkpatrick, J.; Grozema, F.; Andrienko, D.; Kremer, K.; Müllen, K., *Nature Materials* **2009,** *8*, 421.

58. Marcus, R. A., *Reviews of Modern Physics* **1993,** *65* (3), 599.

59. Feng, X.; Wu, J.; Ai, M.; Pisula, W.; Zhi, L.; Rabe, J. P.; Müllen, K., *Angewandte Chemie* **2007,** *119* (17), 3093-3096.

60. Wheeler, S. E., *CrystEngComm* **2012,** *14* (19), 6140-6145.

61. Andrienko, D.; Marcon, V.; Kremer, K., *The Journal of chemical physics* **2006,** *125* (12), 124902.

62. Andrienko, D.; Kirkpatrick, J.; Marcon, V.; Nelson, J.; Kremer, K., *physica status solidi (b)* **2008,** *245* (5), 830-834.

63. Marcon, V.; Vehoff, T.; Kirkpatrick, J.; Jeong, C.; Yoon, D. Y.; Kremer, K.; Andrienko, D., *The Journal of chemical physics* **2008,** *129* (9), 094505.

64. Kirkpatrick, J.; Marcon, V.; Kremer, K.; Nelson, J.; Andrienko, D., *The Journal of chemical physics* **2008,** *129* (9), 094506.

65. Bhola, R.; Bally, T.; Valente, A.; Cyrański, M. K.; Dobrzycki, Ł.; Spain, S. M.; Rempała, P.; Chin, M. R.; King, B. T., *Angewandte Chemie International Edition* **2010,** *49* (2), 399-402.

66. Sakurai, H.; Daiko, T.; Hirao, T., *Science* **2003,** *301* (5641), 1878-1878.

67. Xiao, S.; Myers, M.; Miao, Q.; Sanaur, S.; Pang, K.; Steigerwald, M. L.; Nuckolls, C., *Angewandte Chemie* **2005,** *117* (45), 7556-7560.

68. Luo, J.; Xu, X.; Mao, R.; Miao, Q., *Journal of the American Chemical Society* **2012,** *134* (33), 13796-13803.

69. Shi, K.; Lei, T.; Wang, X.-Y.; Wang, J.-Y.; Pei, J., *Chemical Science* **2014,** *5* (3), 1041-1045.

70. Chen, Q.; Chen, T.; Pan, G.-B.; Yan, H.-J.; Song, W.-G.; Wan, L.-J.; Li, Z.-T.; Wang, Z.-H.; Shang, B.; Yuan, L.-F., *Proceedings of the National Academy of Sciences* **2008,** *105* (44), 16849-16854.

71. Wu, T. C.; Chen, M. K.; Lee, Y. W.; Kuo, M. Y.; Wu, Y. T., *Angewandte Chemie International Edition* **2013,** *52* (4), 1289-1293.

72. Wu, Y.-T.; Siegel, J. S., *Chemical Reviews* **2006,** *106* (12), 4843-4867.

73. Wu, Y.-T.; Bandera, D.; Maag, R.; Linden, A.; Baldridge, K. K.; Siegel, J. S., *Journal of the American Chemical Society* **2008,** *130* (32), 10729-10739.

74. Filatov, A. S.; Petrukhina, M. A., *Journal of Organometallic Chemistry* **2008,** *693* (8), 1590-1596.

75. Wang, B.-T.; Petrukhina, M. A.; Margine, E. R., *Carbon* **2015,** *94*, 174-180.

76. Chiu, C.-Y.; Kim, B.; Gorodetsky, A. A.; Sattler, W.; Wei, S.; Sattler, A.; Steigerwald, M.; Nuckolls, C., *Chemical Science* **2011,** *2* (8), 1480-1486.

77. Pascal, R. A., *Chem. Rev.* **2006,** *106* (12), 4809-4819.

78. Li, Y.; Jia, Z.; Xiao, S.; Liu, H.; Li, Y., *Nature communications* **2016,** *7*, 11637.

79. Houk, K. N.; Cheong, P. H.-Y., *Nature* **2008,** *455* (7211), 309.

80. Houk, K. N.; Liu, F., *Accounts of Chemical Research* **2017,** *50*, 539-543.

81. Hopmann, K. H., *Int. J. Quant. Chem.* **2015,** *115*, 1232-1249.

82. Lam, Y.; Grayson, M. N.; Holland, M. C.; Simon, A.; Houk, K. N., *Accounts of Chemical Research* **2016,** *49*, 750-762.

83. Maji, R.; Mallojjala, S. C.; Wheeler, S. E., *Chemical Society Reviews* **2018**.

84. Jindal, G.; Sunoj, R. B., *Organic & Biomolecular Chemistry* **2014,** *12*, 2745-2753.

85. Seguin, T. J.; Wheeler, S. E., *Angewandte Chemie International Edition* **2016,** *55* (51), 15889-15893.

86. Straker, R. N.; Peng, Q.; Mekareeya, A.; Paton, R. S.; Anderson, E. A., *Nat. Commun.* **2016,** *7*, 10109.

87. 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., *Science* **2017,** *356* (6336), 426-430.

88. Guan, Y.; Wheeler, S. E., *Angewandte Chemie International Edition* **2017,** *56* (31), 9101-9105.

89. Zou, Y.; Gutierrez, O.; Sader, A. C.; Patel, N. D.; Fandrick, D. R.; Busacca, C. A.; Fandrick, K. R.; Kozlowski, M.; Senanayake, C. H., *Organic Letters* **2017**.

90. Schlegel, H. B., *Journal of Computational Chemistry* **1982,** *3* (2), 214-218.

91. Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J., *Journal of Computational Chemistry* **1996,** *17* (1), 49-56.

92. Peng, C.; Schlegel, H. B., *Israel J. Chem.* **1993,** *33*, 449-454.

93. Irikura, K. K.; Johnson, R. D., *The Journal of Physical Chemistry A* **2000,** *104* (11), 2191-2194.

94. Zimmerman, P. M., *Journal of Computational Chemistry* **2015,** *36* (9), 601-611.

95. Varela, J. A.; Vázquez, S. A.; Martínez-Núñez, E., *Chemical science* **2017,** *8* (5), 3843-3851.

96. Maeda, S.; Ohno, K., *Chemical Physics Letters* **2008,** *460* (1-3), 55-58.

97. Ohno, K.; Maeda, S., *The Journal of Physical Chemistry A* **2006,** *110* (28), 8933-8941.

98. Schlegel, H. B., *Theoretica chimica acta* **1992,** *83* (1-2), 15-20.

99. Bondensgård, K.; Jensen, F., *The Journal of chemical physics* **1996,** *104* (20), 8025-8031.

100. Dallos, M.; Lischka, H.; Ventura Do Monte, E.; Hirsch, M.; Quapp, W., *Journal of Computational Chemistry* **2002,** *23* (5), 576-583.

101. Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A., *The Journal of chemical physics* **2004,** *120* (17), 7877-7886.

102. Zimmerman, P., *Journal of chemical theory and computation* **2013,** *9* (7), 3043-3050.

103. Sheppard, D.; Terrell, R.; Henkelman, G., *The Journal of chemical physics* **2008,** *128* (13), 134106.

104. Henkelman, G.; Uberuaga, B. P.; Jónsson, H., *The Journal of chemical physics* **2000,** *113* (22), 9901-9904.

105. Goodrow, A.; Bell, A. T.; Head-Gordon, M., *The Journal of chemical physics* **2008,** *129* (17), 174109.

106. Maeda, S.; Ohno, K.; Morokuma, K., *Physical Chemistry Chemical Physics* **2013,** *15* (11), 3683-3701.

107. Bhoorasingh, P. L.; Slakman, B. L.; Seyedzadeh Khanshan, F.; Cain, J. Y.; West, R. H., *The Journal of Physical Chemistry A* **2017,** *121* (37), 6896-6904.

108. Bhoorasingh, P. L.; West, R. H., *Physical Chemistry Chemical Physics* **2015,** *17* (48), 32173-32182.

109. 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., *International Journal of Quantum Chemistry* **2013,** *113* (18), 2110-2142.

110. Duarte, F.; Paton, R. S., *Journal of the American Chemical Society* **2017,** *139* (26), 8886-8896.

111. Medvedev, M. G.; Panova, M. V.; Chilov, G. G.; Bushmarinov, I. S.; Novikov, F. N.; Stroganov, O. V.; Zeifman, A. A.; Svitanko, I. V., *Mendeleev Communications* **2017,** *27* (3), 224-227.

112. Ardkhean, R.; Roth, P. M.; Maksymowicz, R. M.; Curran, A.; Peng, Q.; Paton, R. S.; Fletcher, S. P., *ACS Catalysis* **2017,** *7* (10), 6729-6737.

113. 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, 2017. https://github.com/ComputationalCatalystDesign/Aaron

114. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., *Journal of cheminformatics* **2012,** *4* (1), 17.

115. Ioannidis, E. I.; Gani, T. Z.; Kulik, H. J., *Journal of Computational Chemistry* **2016,** *37* (22), 2106-2117.

116. RDKit: Open-Source Cheminformatics Software. http://www.rdkit.org/

117. Currently, AaronTools is primarily geared toward Gaussian 09, but is being extended to Gaussian 16 and Orca 4.0.

118. Guan, Y.; Wheeler, S. E., *The Journal of Physical Chemistry C* **2017,** *121* (15), 8541-8547.

119. Although the focus here is on stereoselectivity, AARON is perfectly suitable for looking at regioselectivity, etc.

120. Lu, G.; Fang, C.; Xu, T.; Dong, G.; Liu, P., *Journal of the American Chemical Society* **2015,** *137* (25), 8274-8283.

121. Wang, S.-G.; Liu, X.-J.; Zhao, Q.-C.; Zheng, C.; Wang, S.-B.; You, S.-L., *Angew. Chem. Int. Ed.* **2015,** *54*, 14929-14932.

122. Hawkins, P. C. D., *Journal of Chemical Information and Modeling* **2017,** *57* (8), 1747-1756.

123. Grimme, S., *Chemistry-A European Journal* **2012,** *18* (32), 9955-9964.

124. Chen, S.; Zheng, Y.; Cui, T.; Meggers, E.; Houk, K. N., *Journal of the American Chemical Society* **2018,** *140* (15), 5146-5152.

125. Werner, E. W.; Mei, T.-S.; Burckle, A. J.; Sigman, M. S., *Science* **2012,** *338* (6113), 1455-1458.

126. Zhang, C.; Santiago, C. B.; Crawford, J. M.; Sigman, M. S., *Journal of the American Chemical Society* **2015,** *137* (50), 15668-15671.

127. Zhang, C.; Santiago, C. B.; Kou, L.; Sigman, M. S., *Journal of the American Chemical Society* **2015,** *137* (23), 7290-7293.

128. Patel, H. H.; Sigman, M. S., *Journal of the American Chemical Society* **2015,** *137* (10), 3462-3465.

129. Mei, T.-S.; Patel, H. H.; Sigman, M. S., *Nature* **2014,** *508* (7496), 340.

130. Xu, L.; Hilton, M. J.; Zhang, X.; Norrby, P.-O.; Wu, Y.-D.; Sigman, M. S.; Wiest, O., *Journal of the American Chemical Society* **2014,** *136* (5), 1960-1967.

131. Hilton, M. J.; Xu, L.-P.; Norrby, P.-O.; Wu, Y.-D.; Wiest, O.; Sigman, M. S., *The Journal of organic chemistry* **2014,** *79* (24), 11841-11850.

132. Zhang, C.; Tutkowski, B.; DeLuca, R. J.; Joyce, L. A.; Wiest, O.; Sigman, M. S., *Chemical science* **2017,** *8* (3), 2277-2282.

133. Landis, C. R.; Hilfenhaus, P.; Feldgus, S., *J. Am. Chem. Soc.* **1999,** *121* (38), 8741-8754.

134. Landis, C. R., *Angew. Chem. Int. Ed.* **2000,** *39* (16), 2863-2866.

135. Feldgus, S.; Landis, C. R., *Organometallics* **2001,** *20* (11), 2374-2386.

136. Donoghue, P. J.; Helquist, P.; Wiest, O., *J. Org. Chem.* **2007,** *72* (3), 839-847.

137. Liu, G.; Liu, X.; Cai, Z.; Jiao, G.; Xu, G.; Tang, W., *Angew. Chem. Int. Ed.* **2013,** *52* (15), 4235-4238.

138. Rooks, B. J.; Haas, M. R.; Sepúlveda, D.; Lu, T.; Wheeler, S. E., *ACS Catalysis* **2014,** *5* (1), 272-280.

139. Doney, A. C.; Rooks, B. J.; Lu, T.; Wheeler, S. E., *ACS Catalysis* **2016,** *6* (11), 7948-7955.

140. Sepúlveda, D.; Lu, T.; Wheeler, S. E., *Organic & biomolecular chemistry* **2014,** *12* (41), 8346-8353.

141. Lu, T.; Zhu, R.; An, Y.; Wheeler, S. E., *Journal of the American Chemical Society* **2012,** *134* (6), 3095-3102.

142. 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.

143. Zimmerman, P. M., *Journal of Computational Chemistry* **2013,** *34* (16), 1385-1392.

144. Pendleton, I. M.; Pérez-Temprano, M. H.; Sanford, M. S.; Zimmerman, P. M., *Journal of the American Chemical Society* **2016,** *138* (18), 6049-6060.

145. Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A., *Journal of chemical theory and computation* **2014,** *10* (3), 897-907.

146. Bergeler, M.; Simm, G. N.; Proppe, J.; Reiher, M., *Journal of chemical theory and computation* **2015,** *11* (12), 5712-5722.

147. Suleimanov, Y. V.; Green, W. H., *Journal of chemical theory and computation* **2015,** *11* (9), 4248-4259.

148. Lee, J.; Lee, I.-H.; Joung, I.; Lee, J.; Brooks, B. R., *Nature Communications* **2017,** *8*, 15443.

149. Houk, K. N.; Cheong, P. H.-Y., *Nature* **2008,** *455*, 309-313.

150. Jindal, G.; Sunoj, R. B., *Org. Biomol. Chem.* **2014,** *12*, 2745-2753.

151. Lam, Y.; Grayson, M. N.; Holland, M. C.; Simon, A.; Houk, K. N., *Acc. Chem. Res.* **2016,** *49*, 750-762.

152. Poree, C.; Schoenbeck, F., *Acc. Chem. Res.* **2017,** *50*, 605-608.

153. Houk, K. N.; Liu, F., *Acc. Chem. Res.* **2017,** *50*, 539-543.

154. Santoro, S.; Kalek, M.; Huang, G.; Himo, F., *Acc. Chem. Res.* **2016,** *49*, 1006-1018.

155. Guan, Y.; Rooks, B. J.; Wheeler, S. E. *AARON: An Automated Reaction Optimizer for New catalysts, v. 0.95, Y. Guan, B. J. Rooks, and S. E. Wheeler, Texas A&M University, College Station, TX, 2017.*, 0.95; Texas A&M University, College Station, TX, 2017.

156. Rooks, B. J.; Haas, M. R.; Sepúlveda, D.; Lu, T.; Wheeler, S. E., *ACS Catal.* **2015,** *5*, 272-280.

157. Doney, A. C.; Rooks, B. J.; Lu, T.; Wheeler, S. E., *ACS Catal.* **2016,** *6*, 7948-7955.

158. For methods to automatically identify TSs for reactions with unknown mechanisms, see a) W. M. C. Sameera, S. Maeda, K. Morokuma, Acc. Chem. Res. 2016, 49, 763-773; b) S. Maeda, K. Ohno, K. Morokuma, Phys. Chem. Chem. Phys. 2013, 15, 3683-3701; c) S. Habershon, J. Chem. Phys., 2015, 143, 094106.; and d) A. L. Dewyer, P. M. Zimmerman, Org. Biomol. Chem. 2017, 15, 501-504.

159. Knowles, W. S., *Acc. Chem. Res.* **1983,** *16* (3), 106-112.

160. Noyori, R., *New York* **1994**.

161. Burk, M. J., *Acc. Chem. Res.* **2000,** *33* (6), 363-372.

162. Tang, W.; Zhang, X., *Chem. Rev.* **2003,** *103* (8), 3029-3070.

163. Zhang, W.; Chi, Y.; Zhang, X., *Acc. Chem. Res.* **2007,** *40* (12), 1278-1290.

164. Johnson, N. B.; Lennon, I. C.; Moran, P. H.; Ramsden, J. A., *Acc. Chem. Res.* **2007,** *40* (12), 1291-1299.

165. Minnaard, A. J.; Feringa, B. L.; Lefort, L.; de Vries, J. G., *Acc. Chem. Res.* **2007,** *40* (12), 1267-1277.

166. Blaser, H.-U.; Brieden, W.; Pugin, B.; Spindler, F.; Studer, M.; Togni, A., *Top. Catal.* **2002,** *19* (1), 3-16.

167. Hallock, Y. F.; Manfredi, K. P.; Dai, J.-R.; Cardellina, J. H.; Gulakowski, R. J.; McMahon, J. B.; Schäffer, M.; Stahl, M.; Gulden, K.-P.; Bringmann, G., *J. Nat. Prod.* **1997,** *60* (7), 677-683.

168. Nugent, T. C., *Chiral Amine Synthesis: Methods, Developments and Applications*. John Wiley & Sons: 2010.

169. Bringmann, G.; Gulder, T.; Gulder, T. A.; Breuning, M., *Chem. Rev.* **2010,** *111* (2), 563-639.

170. Xu, G.; Fu, W.; Liu, G.; Senanayake, C. H.; Tang, W., *J. Am. Chem. Soc.* **2014,** *136* (2), 570-573.

171. Huang, L.; Zhu, J.; Jiao, G.; Wang, Z.; Yu, X.; Deng, W. P.; Tang, W., *Angew. Chem. Int. Ed.* **2016,** *55*, 4527-4531.

172. Zhao, Y.; Truhlar, D. G., *J. Chem. Phys.* **2006,** *125* (19), 194101.

173. Hay, P. J.; Wadt, W. R., *J. Chem. Phys.* **1985,** *82* (1), 299-310.

174. Yang, Y.; Weaver, M. N.; Merz, K. M., *J. Phys. Chem. A* **2009,** *113* (36), 9843-9851.

175. Ye, L.; Wang, Y.; Aue, D. H.; Zhang, L., *J. Am. Chem. Soc.* **2012,** *134* (1), 31-34.

176. Sarwar, M. G.; Dragisic, B.; Salsberg, L. J.; Gouliaras, C.; Taylor, M. S., *J. Am. Chem. Soc.* **2010,** *132* (5), 1646-1653.

177. Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.* **2008,** *120* (1), 215-241.

178. Andrae, D.; Haeußermann, U.; Dolg, M.; Stoll, H.; Preuss, H., *Theor. Chem. Acc.* **1990,** *77*, 123-141.

179. Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H., *J. Chem. Phys.* **1987,** *86*, 866-872.

180. Miertus, S.; Scrocco, E.; Tomasi, J., *Chem. Phys.* **1981,** *55* (1), 117-129.

181. Tomasi, J.; Mennucci, B.; Cammi, R., *Chem. Rev.* **2005,** *105* (8), 2999-3093.

182. Burk, M. J.; Wang, Y. M.; Lee, J. R., *J. Am. Chem. Soc.* **1996,** *118* (21), 5142-5143.

183. Chen, J.; Zhang, W.; Geng, H.; Li, W.; Hou, G.; Lei, A.; Zhang, X., *Angew. Chem. Int. Ed.* **2009,** *48* (4), 800-802.

184. Halpern, J., *Science* **1982,** *217* (4558), 401-407.

185. Landis, C. R.; Halpern, J., *J. Am. Chem. Soc* **1987,** *109* (6), 1746-1754.

186. Feldgus, S.; Landis, C. R., *J. Am. Chem. Soc.* **2000,** *122* (51), 12714-12727.

187. Gridnev, I. D.; Yasutake, M.; Higashi, N.; Imamoto, T., *J. Am. Chem. Soc.* **2001,** *123* (22), 5268-5276.

188. Imamoto, T.; Tamura, K.; Zhang, Z.; Horiuchi, Y.; Sugiya, M.; Yoshida, K.; Yanagisawa, A.; Gridnev, I. D., *J. Am. Chem. Soc.* **2012,** *134* (3), 1754-1769.

189. Yasutake, M.; Gridnev, I. D.; Higashi, N.; Imamoto, T., *Org. Lett.* **2001,** *3* (11), 1701-1704.

190. Reetz, M. T.; Meiswinkel, A.; Mehler, G.; Angermund, K.; Graf, M.; Thiel, W.; Mynott, R.; Blackmond, D. G., *J. Am. Chem. Soc.* **2005,** *127* (29), 10305-10313.

191. Imamoto, T.; Yashio, K.; Crépy, K. V.; Katagiri, K.; Takahashi, H.; Kouchi, M.; Gridnev, I. D., *Organometallics* **2006,** *25* (4), 908-914.

192. Gridnev, I. D.; Kohrt, C.; Liu, Y., *Dalton Trans.* **2014,** *43* (4), 1785-1790.

193. Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O., *J. Am. Chem. Soc.* **2008,** *131* (2), 410-411.

194. Gridnev, I. D.; Imamoto, T., *Chem. Commun.* **2009,** (48), 7447-7464.

195. Eksterowicz, J. E.; Houk, K. N., *Chem. Rev.* **1993,** *93*, 2439-2461.

196. Corbeil, C. R.; Thielges, S.; Schwartzentruber, J. A.; Moitessier, N., *Angew. Chem. Int. Ed.* **2008,** *47*, 2635-2638.

197. Jensen, F., *J. Chem. Phys.* **2003,** *119*, 8804-8808.

198. Madarász, Á.; Berta, D.; Paton, R. S., *J. Chem. Theory Comput.* **2016,** *12*, 1833-1844.

199. Since TS1αY and TS1βY involve an initial hydride transfer to different carbon atoms, TS1αY cannot lead to TS2βY and vice versa.

200. Since TS1αY and TS1βY involve an initial hydride transfer to different carbon atoms, TS1αY cannot lead to TS2βY and vice versa.

201. For example, for L2, considering two orientations for each OMe on the substrate and catalyst leads to 16 conformations for each of the 12 possible transition states.

202. In this case, for L1 the TS library was generated using a simplified version of the ligand in which the two tBu groups were replaced with Me and the OMe groups

removed from the substrate. For subsequent ligands, the TS library comprised the final optimized TS structures for the previous ligand.

203. 393 initial TS structures were generated by AARON, resulting in 251 final optimized structures. Among the 142 unfinished TS searches, a vast majority were terminated because they were converging to conformations that had previously been located. The remainder either failed to converge after multiple attempts or converged to the incorrect transition state.

204. "Catalyst" here is the ligand bound to Rh along with one nbd.

205. Hammes-Schiffer, S., *Acc. Chem. Res.* **2017,** *50*, 561-566.

206. Sakurai, H.; Daiko, T.; Sakane, H.; Amaya, T.; Hirao, T., *J. Am. Chem. Soc.* **2005,** *127* (33), 11580-11581.

207. Amaya, T.; Hirao, T., *Chem. Commun.* **2011,** *47* (38), 10524-10535.

208. Wu, Y.-T.; Siegel, J. S., *Chem. Rev.* **2006,** *106* (12), 4843-4867.

209. Amaya, T.; Seki, S.; Moriuchi, T.; Nakamoto, K.; Nakata, T.; Sakane, H.; Saeki, A.; Tagawa, S.; Hirao, T., *J. Am. Chem. Soc.* **2008,** *131* (2), 408-409.

210. Feng, X.; Marcon, V.; Pisula, W.; Hansen, M. R.; Kirkpatrick, J.; Grozema, F.; Andrienko, D.; Kremer, K.; Müllen, K., *Nat. Mater.* **2009,** *8* (5), 421-426.

211. Zoppi, L.; Martin-Samos, L.; Baldridge, K. K., *J. Am. Chem. Soc.* **2011,** *133*, 14002-14009.

212. Wang, B.-T.; Petrukhina, M. A.; Margine, E. R., *Carbon* **2015,** *94*, 174-180.

213. Wu, Y.-T.; Bandera, D.; Maag, R.; Linden, A.; Baldridge, K. K.; Siegel, J. S., *J. Am. Chem. Soc.* **2008,** *130*, 10729-10739.

214. Amaya, T.; Hifumi, M.; Okada, M.; Shimizu, Y.; Moriuchi, T.; Segawa, K.; Ando, Y.; Hirao, T., *J. Org. Chem.* **2011,** *76* (19), 8049-8052.

215. Schmidt, B. M.; Topolinski, B.; Higashibayashi, S.; Kojima, T.; Kawano, M.; Lentz, D.; Sakurai, H., *Chem. Eur. J.* **2013,** *19* (10), 3282-3286.

216. Higashibayashi, S.; Tsuruoka, R.; Soujanya, Y.; Purushotham, U.; Sastry, G. N.; Seki, S.; Ishikawa, T.; Toyota, S.; Sakurai, H., *Bull. Chem. Soc. Jpn.* **2012,** *85* (4), 450-467.

217. Nishio, M., *CrystEngComm* **2004,** *6* (27), 130-158.

218. Shrestha, B. B.; Morita, Y.; Kojima, T.; Kawano, M.; Higashibayashi, S.; Sakurai, H., *Chem. Lett.* **2014,** *43* (8), 1294-1296.

219. Chen, X.; Bai, F.-Q.; Tang, Y.; Zhang, H.-X., *J. Comput. Chem.* **2015,** *37*, 813-824.

220. Thorley, K. J.; Finn, T. W.; Jarolimek, K.; Anthony, J. E.; Risko, C., *Chem. Mater.* **2017,** *Article ASAP, DOI: 10.1021/acs.chemmater.6b04211*.

221. Hunter, C. A.; Sanders, J. K. M., *J. Am. Chem. Soc.* **1990,** *112*, 5525-5534.

222. Raju, R. K.; Bloom, J. W. G.; An, Y.; Wheeler, S. E., *ChemPhysChem* **2011,** *12*, 3116-3130.

223. Wheeler, S. E., *Acc. Chem. Res.* **2013,** *46*, 1029-1038.

224. Wheeler, S. E.; Bloom, J. W. G., *J. Phys. Chem. A* **2014,** *118*, 6133-6147.

225. Sinnokrot, M. O.; Sherrill, C. D., *J. Phys. Chem. A* **2006,** *110*, 10656-10668.

226. Kennedy, M. R.; Burns, L. A.; Sherrill, C. D., *J. Phys. Chem. A* **2012,** *116*, 11920-11926.

227. Chai, J.-D.; Head-Gordon, M., *J. Chem. Phys.* **2008,** *128*, 084106.

228. Schäfer, A.; Huber, C.; Ahlrichs, R., *J. Chem. Phys.* **1994,** *100* (8), 5829-5835.

229. Papajak, E.; Truhlar, D. G., *J. Chem. Theory Comput.* **2011,** *7*, 10-18.

230. Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D., *J. Chem. Phys.* **2014,** *140*, 094106.

231. Jorgensen, W. L.; Madura, J. D.; Swenson, C. J., *J. Am. Chem. Soc.* **1984,** *106* (22), 6638-6646.

232. Jorgensen, W. L.; Severance, D. L., *J. Am. Chem. Soc.* **1990,** *112* (12), 4768-4774.

233. Jorgensen, W. L., OPLS Force Fields. In *Encyclopedia of Computational Chemistry*, John Wiley & Sons, Ltd: 2002.

234. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.

235. Singh, U. C.; Kollman, P. A., *J. Comput. Chem.* **1984,** *5* (2), 129-145.

236. Besler, B. H.; Merz, K. M.; Kollman, P. A., *J. Comput. Chem.* **1990,** *11* (4), 431-439.

237. Becke, A. D., *Phys. Rev. A* **1988,** *38* (6), 3098.

238. Lee, C.; Yang, W.; Parr, R. G., *Physical review B* **1988,** *37* (2), 785.

239. Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H., *Chem. Phys. Lett.* **1989,** *157* (3), 200-206.

240. Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J., *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.

241. Nosé, S., *Mol. Phys.* **1984,** *52* (2), 255-268.

242. Hoover, W. G., *Phys. Rev. A* **1985,** *31* (3), 1695.

243. Parrinello, M.; Rahman, A., *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.

244. Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D., *WIREs Comp. Mol. Sci.* **2012,** *2*, 556-565.

245. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E., *J. Chem. Theory Comput.* **2008,** *4* (3), 435-447.

246. Grimme, S., *J. Comput. Chem.* **2006,** *27* (15), 1787-1799.

247. Szalewicz, K., *WIREs Comp. Mol. Sci.* **2012,** *2*, 254-272.

248. Jeziorski, B.; Moszyński, R.; Szalewicz, K., *Chem. Rev.* **1994,** *94*, 1887-1930.

249. Hohenstein, E. G.; Sherrill, C. D., *WIREs Comp. Mol. Sci.* **2012,** *2*, 304-326.

250. The total SAPT0 interaction energies are plotted in Appendix B Figure B-6, and closely resemble the B97-D data. Induction energies are also plotted in Figure B-7, and vary by less than 1 kcal mol-1 across all angles.

251. At equal intermonomer distances of 3.81 A, the exchange repulsion interaction is 3 kcal mol-1 lower for the stacked dimer of 4 than for 1.

252. For 2 and 3, which lack C3-symmetry, angles from 0° to 180° between CO/CH2 groups were considered (see Appendix B Figure B-5). Local orientations of 60° and

120° correspond to energy minima. For 2, 60° is the global minimum while for 3, the energy at 120° is lower than at 60° by ~0.2 kcal mol-1. This difference can be attributed to stronger exchange-repulsion for 3 at 60°.

253. Because 2 and 3 lack C3-symmetry, consideration of angles up to 360° are necessary (see Appendix B Figure B-12). For 2, there are peaks at 60° and 300° in the angle distribution, which is consistent with the X-ray analysis of 60° stacking. For 3, there are two major peaks near 180°, seemingly at odds with the dimer interaction energies where 120° is the global minimum.

254. Johnston, R. C.; Cheong, P. H.-Y., *Org. Biomol. Chem.* **2013,** *11*, 5057-5064.

255. Filatov, A.; Petrukhina, M. A., *J. Organomet. Chem.* **2008,** *693*, 1590-1596.

256. Forrest, S. R., *Nature* **2004,** *428* (6986), 911-918.

257. Pisula, W.; Menon, A.; Stepputat, M.; Lieberwirth, I.; Kolb, U.; Tracz, A.; Sirringhaus, H.; Pakula, T.; Müllen, K., *Adv. Mater. (Weinheim, Ger.)* **2005,** *17* (6), 684-689.

258. Ball, M.; Zhong, Y.; Wu, Y.; Schenck, C.; Ng, F.; Steigerwald, M.; Xiao, S.; Nuckolls, C., *Acc. Chem. Res.* **2015,** *48* (2), 267-276.

259. Wheeler, S. E., *J. Am. Chem. Soc.* **2011,** *133* (26), 10262-10274.

260. Wheeler, S. E., *Acc. Chem. Res.* **2013,** *46* (4), 1029-1038.

261. Wheeler, S. E.; Bloom, J. W. G., *J. Phys. Chem. A* **2014,** *118* (32), 6133-6147.

262. Grimme, S., *Angew. Chem. Int. Ed.* **2008,** *47*, 3430-3434.

263. Martinez, C. R.; Iverson, B. L., *Chem. Sci.* **2012,** *3*, 2191-2201.

264. Bloom, J. W. G.; Wheeler, S. E., *Angew. Chem. Int. Ed.* **2011,** *50*, 7847-7849.

265. Herein, we use the term "stacking" purely as a geometric descriptor, and avoid the use of "p-stacking" or "p-p stacking" to avoid implying any special feature of stacking interactions involving aromatic rings. See Refs. 9-11.

266. Haddon, R., *Science* **1993,** *261* (5128), 1545-1550.

267. Haddon, R., *Acc. Chem. Res.* **1992,** *25* (3), 127-133.

268. Kennedy, M. R.; Burns, L. A.; Sherrill, C. D., *J. Phys. Chem. A* **2012,** *116* (48), 11920-11926.

269. Cui, Z.-h.; Gupta, A.; Lischka, H.; Kertesz, M., *Phys. Chem. Chem. Phys.* **2015,** *17*, 23963-23969.

270. Xiao, S.; Myers, M.; Miao, Q.; Sanaur, S.; Pang, K.; Steigerwald, M. L.; Nuckolls, C., *Angew. Chem. Int. Ed.* **2005,** *44* (45), 7390–7394.

271. Schiros, T.; Mannsfeld, S.; Chiu, C. y.; Yager, K. G.; Ciston, J.; Gorodetsky, A. A.; Palma, M.; Bullard, Z.; Kramer, T.; Delongchamp, D., *Adv. Funct. Mater.* **2012,** *22* (6), 1167-1173.

272. Chen, L.; Puniredd, S. R.; Tan, Y.-Z.; Baumgarten, M.; Zschieschang, U.; Enkelmann, V.; Pisula, W.; Feng, X.; Klauk, H.; Müllen, K., *J. Am. Chem. Soc.* **2012,** *134* (43), 17869-17872.

273. Chen, L.; Mali, K. S.; Puniredd, S. R.; Baumgarten, M.; Parvez, K.; Pisula, W.; De Feyter, S.; Müllen, K., *J. Am. Chem. Soc.* **2013,** *135* (36), 13531-13537.

274. Wang, X.-Y.; Zhuang, F.-D.; Wang, R.-B.; Wang, X.-C.; Cao, X.-Y.; Wang, J.-Y.; Pei, J., *J. Am. Chem. Soc.* **2014,** *136* (10), 3764-3767.

275. Wu, Y.-T.; Bandera, D.; Maag, R.; Linden, A.; Baldridge, K. K.; Siegel, J. S., *J. Am. Chem. Soc.* **2008,** *130*, 10729-10739.

276. Caetano, E. W.; Freire, V. N.; Dos Santos, S.; Galvao, D. S.; Sato, F., *J. Chem. Phys.* **2008,** *128* (16), 164719.

277. Zhong, Y.; Kumar, B.; Oh, S.; Trinh, M. T.; Wu, Y.; Elbert, K.; Li, P.; Zhu, X.; Xiao, S.; Ng, F.; Steigerwald, M. L.; Nuckolls, C., *J. Am. Chem. Soc.* **2014,** *136* (22), 8122-8130.

278. Li, Y.; Jia, Z.; Xiao, S.; Liu, H.; Li, Y., *Nature communications* **2016,** *7*.

279. Chalifoux, W. A., *Angew. Chem. Int. Ed.* **2017,** *Early View (DOI: 10.1002/anie.201702687)*.

280. Daigle, M.; Miao, D.; Lucotti, A.; Tommasini, M., *Angew. Chem. Int. Ed.* **2017,** *Early View (DOI: 10.1002/anie.201611834)*.

281. Bheemireddy, S. R.; Ubaldo, P. C.; Finke, A. D.; Wang, L.; Plunkett, K. N., *J. Mater. Chem. C* **2016,** *4*, 3963-3969.

282. Whalley, A. C.; Plunkett, K. N.; Gorodetsky, A. A.; Schenck, C.; Chiu, C.-Y.; Steigerwald, M.; Nuckolls, C., *Chem. Sci.* **2011,** *2*, 132-135.

283. Kang, S. J.; Kim, J. B.; Chiu, C.-Y.; Ahn, S.; Schiros, T.; Lee, S. S.; Yager, K. G.; Toney, M. F.; Loo, Y.-L.; Nuckolls, C., *Angew. Chem. Int. Ed.* **2012,** *51* (34), 8594-8597.

284. Schiros, T.; Kladnik, G.; Prezzi, D.; Ferretti, A.; Olivieri, G.; Cossaro, A.; Floreano, L.; Verdini, A.; Schenck, C.; Cox, M.; Gorodetsky, A. A.; Plunkett, K. N.;

Delongchamp, D.; Nuckolls, C.; Morgante, A.; Cvetko, D.; Kymissis, *Adv. Energy Mater.* **2013,** *3*, 894-902.

285. Guan, Y.; Wheeler, S. E., *J. Phys. Chem. C* **2017**.

286. Plunkett, K. N.; Godula, K.; Nuckolls, C.; Tremblay, N.; Whalley, A. C.; Xiao, S., *Org. Lett.* **2009,** *11*, 2225-2228.

287. Yamamoto, K.; Harada, T.; Okamoto, Y.; Chikamatsu, H.; Nakazaki, M.; Kai, Y.; Nakao, T.; Tanaka, M.; Harada, S.; Kasai, N., *J. Am. Chem. Soc.* **1988,** *110* (11), 3578-3584.

288. Yamamoto, K.; Harada, T.; Nakazaki, M.; Naka, T.; Kai, Y.; Harada, S.; Kasai, N., *J. Am. Chem. Soc.* **1983,** *105* (24), 7171-7172.

289. Feng, C.-N.; Kuo, M.-Y.; Wu, Y.-T., *Angew. Chem. Int. Ed.* **2013,** *52* (30), 7791-7794.

290. Miller, R. W.; Duncan, A. K.; Schneebeli, S. T.; Gray, D. L.; Whalley, A. C., *Chem. Eur. J.* **2014,** *20* (13), 3705-3711.

291. Hiszpanski, A. M.; Baur, R. M.; Kim, B.; Tremblay, N. J.; Nuckolls, C.; Woll, A. R.; Loo, Y.-L., *J. Am. Chem. Soc.* **2014,** *136* (44), 15749-15756.

292. Christoph, H.; Grunenberg, J.; Hopf, H.; Dix, I.; Jones, P. G.; Scholtissek, M.; Maier, G., *Chem. Eur. J.* **2008,** *14* (18), 5604-5616.

293. Sakamoto, Y.; Suzuki, T., *J. Am. Chem. Soc.* **2013,** *135* (38), 14074-14077.

294. Salcedo, R.; Sansores, L. E.; Picazo, A.; Sansón, L., *J. Mol. Struct.: THEOCHEM* **2004,** *678* (1), 211-215.

295. Corbeil, C. R.; Thielges, S.; Schwartzentruber, J. A.; Moitessier, N., *Angewandte Chemie International Edition* **2008,** *47* (14), 2635-2638.

296. Becke, A., *J. Chem. Phys.* **1997,** *107*, 8554-8560.

297. Grimme, S., *J. Comp. Chem.* **2006,** *27*, 1787-1799.

298. Schafer, A.; Huber, C.; Ahlrichs, R., *J. Chem. Phys.* **1994,** *100*, 5829-5835.

299. Dunning, T. H., Jr., *J. Chem. Phys.* **1989,** *90*, 1007-1023.

300. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., *J. Chem. Phys.* **2010,** *132* (15), 154104.

301. Janowski, T.; Pulay, P., *J. Am. Chem. Soc.* **2012,** *134* (42), 17520-17525.

302. Grimme, S.; Ehrlich, S.; Goerigk, L., *J. Comput. Chem.* **2011,** *32* (7), 1456-1465.

303. Smith, D. G.; Burns, L. A.; Patkowski, K.; Sherrill, C. D., *J. Chem. Phys.* **2016,** *7* (12), 2197-2203.

304. Marcon, V.; Vehoff, T.; Kirkpatrick, J.; Jeong, C.; Yoon, D. Y.; Kremer, K.; Andrienko, D., *J. Chem. Phys.* **2008,** *129* (9), 094505.

305. Sinnokrot, M. O.; Sherrill, C. D., *J. Phys. Chem. A* **2004,** *108* (46), 10200-10207.

306. In Table 2, the binding energies of the Saddle-like dimers of 2 and 3 are equal. However, this is largely becuase the S conformation is relatively high in energy for 2, but not for 3. If the binding energies for these two dimers are computed relative to the corresponding monomer conformation, the binding energies are 44.9 and 43.9 kcal/mol for 2 (S2) and 3 (S1), repectively.

307. Rodríguez-Ropero, F.; Casanovas, J.; Alemán, C., *J. Comput. Chem.* **2008,** *29* (1), 69-78.

308. The elephant seal configuration is named due to the uncanny resemblance of this configuration to the posturing of battling male elephant seals.

309. Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D., *J. Am. Chem. Soc.* **2002,** *124*, 10887-10893.

310. Wheeler, S. E.; Houk, K. N., *Mol. Phys.* **2009,** *107*, 749-760.

APPENDIX A

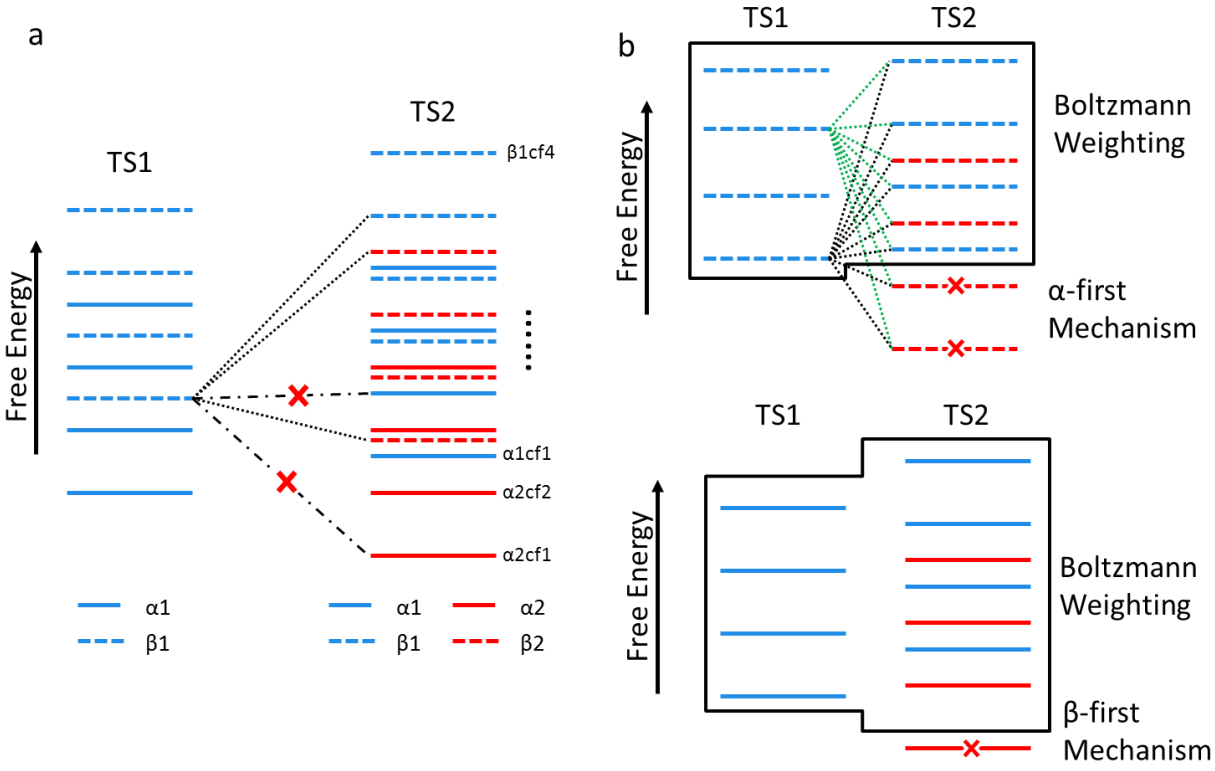


Figure A-1. Boltzmann weighting for multiple reaction pathways.

In this series of reactions, there are many thermodynamically accessible pathways leading to both the major and minor stereoisomers, and we assume rapid interconversion between intermediates that lie between TS1 and TS2 (as long as the intermediates correspond to the same mechanism and same stereoisomer). The possible different pathways were identified as follows. Figure A-1(a) shows hypothetical relative free energies for TS1 and TS2. Solid lines represent TS structures corresponding to α-carbon first pathways, while dash lines represent the β-carbon first TS structures. Following traversal of TS1, each pathway can pass through at least two potential

configurations of TS2. For TS2, blue lines indicate TS2α1 or TS2β1, while red lines indicate TS2α2 and TS2β2. Connection between TS1 and TS2 exist only for lines with the same type (both solid or both dashed). In this way, all possible pathways can be enumerated.

For the Boltzmann weighting over these pathways, we consider the α-carbon first and β-carbon first mechanisms separately (Figure A-1b). For the pathways for a given mechanism, the reaction can pass through any combination of the different configurations for TS1 and TS2. In cases in which the second step is lower in free energy than the first step, that second step (lines marked with a red "X" in Figure A-1b) can never be stereodetermining, while all other transition states impact the overall rate of formation of a given stereoisomer (transition states within the black border in Figure A-1b). Performing a Boltzmann weighting over these TS structures and summing over the two possible mechanisms (α-first and β-first) we can get the Boltzmann weighted *ee*'s by the following equation,

$$ee = \frac{\sum_M \sum_{ts} e^{\Delta\Delta G(R)/RT} - \sum_M \sum_{ts} e^{\Delta\Delta G(S)/RT}}{\sum_M \sum_{ts} e^{\Delta\Delta G(R)/RT} + \sum_M \sum_{ts} e^{\Delta\Delta G(S)/RT}}$$

in which ΔΔG(R) and ΔΔG(S) are free energy barriers for transition states relative to the lowest-lying transition state for R and S enantiomers, respectively. Σ_{ts} represents a sum over all the transition states as described above, Σ_M represents a sum over the α-carbon and β-carbon first the mechanisms.

APPENDIX B

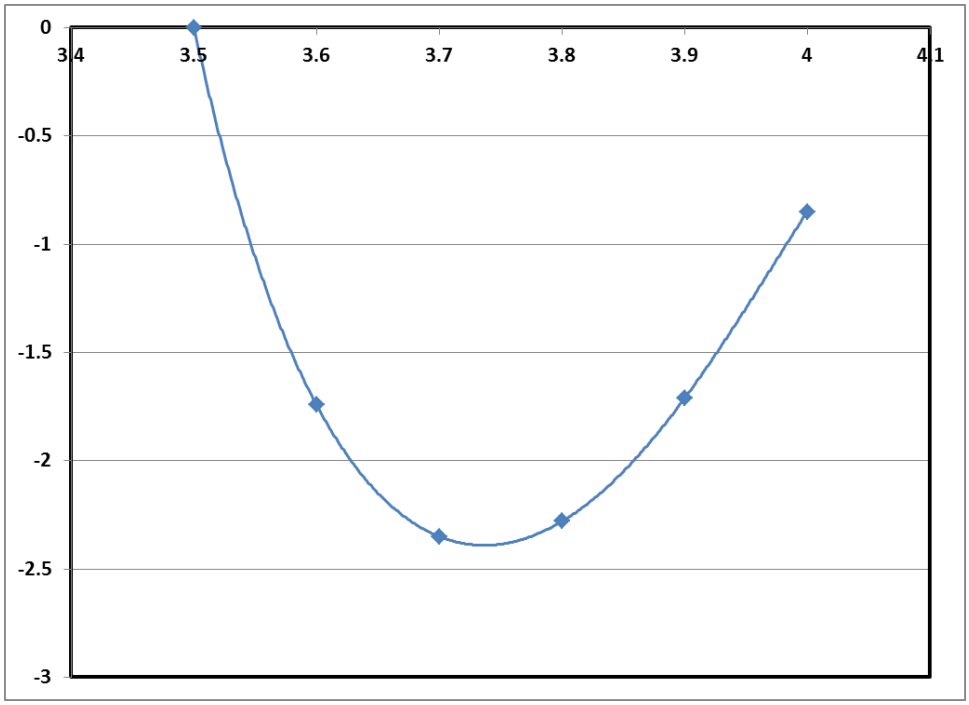


Figure B-1. Relative energy of sumanenetrione at $\phi = 0°$ as the function of bowl-to-bowl distance and the fitting of a 5^{th} order polynomial.

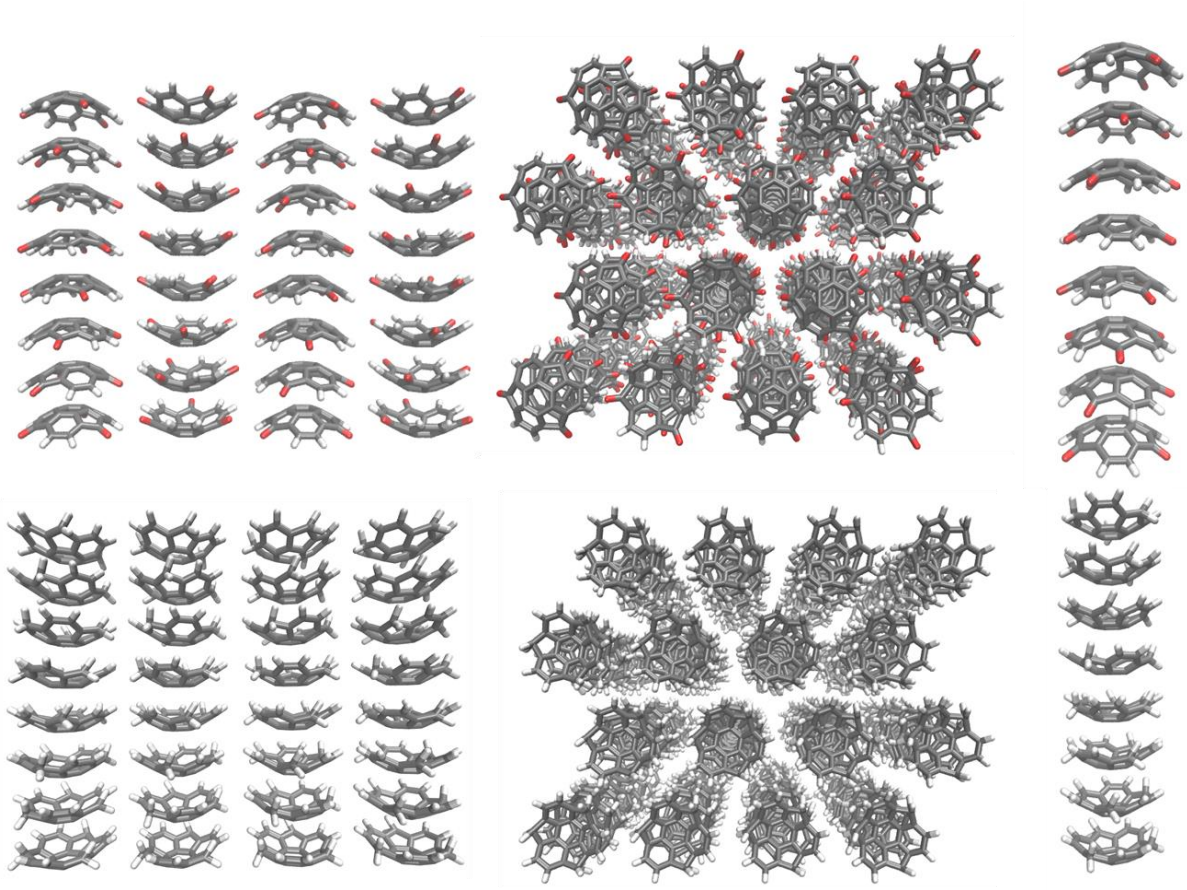


Figure B-2. Initial geometries for MD simulations of sumanenetrione 4 (top) and sumanene 1 (bottom). 128 molecules (4*4*8) were input for each molecule. Left: side view for four columns (for molecule 2-4 neighboring columns are in opposite direction; for molecule 1 columns are in unidirection.); Middle: top view; Right: side view of single column (30° orientation between neighboring molecules).

To make sure that the system has reached an equilium, the orientation angle distribution between neighboring molecules was monitored for 10 ns after the production run.

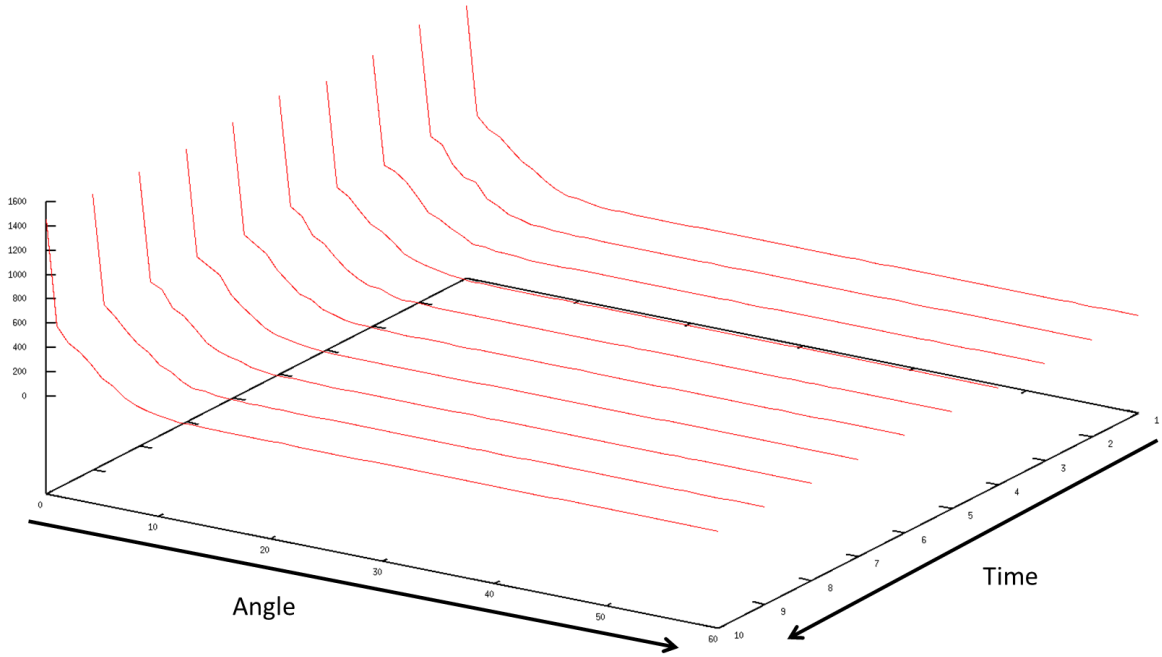


Figure B-3. Angle distribution as a function of time within 10 ns following production run for sumanenetrione 4.

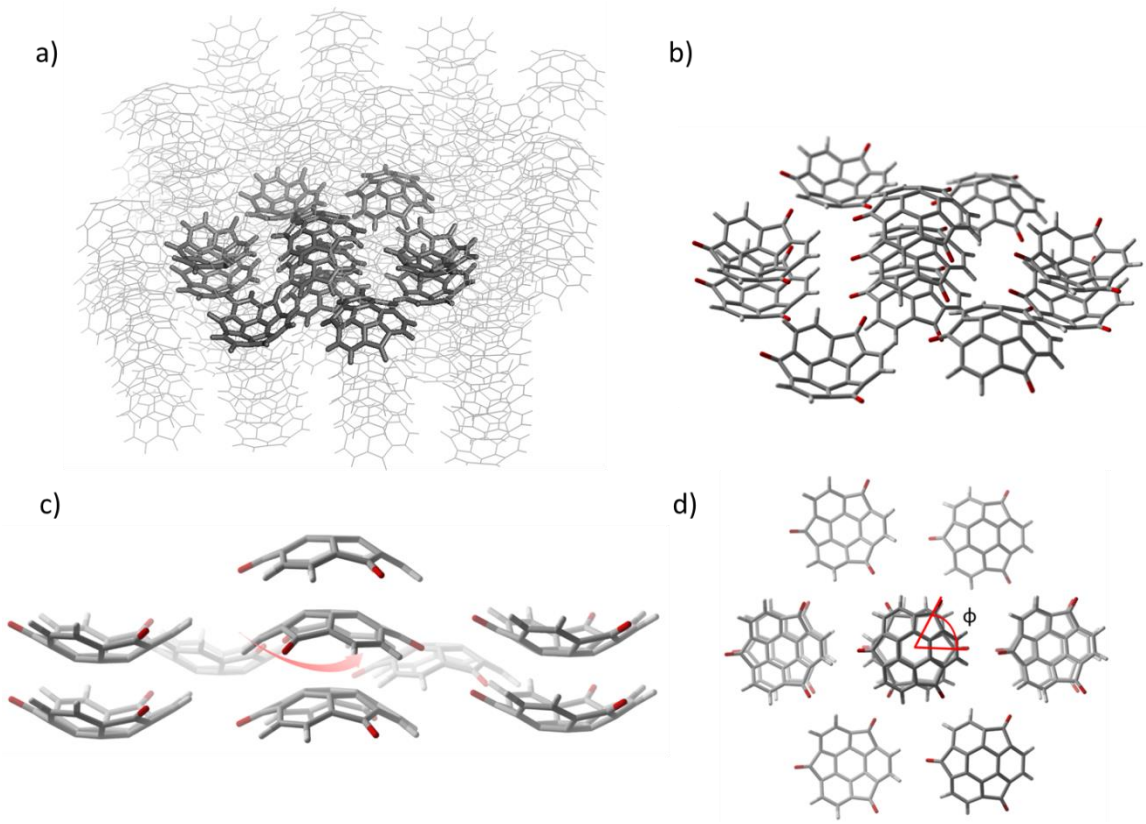


Figure B-4. Construction of a cluster model involving both intracolumnar and intercolumnar interactions a) 11-molecule cluster extracted from the last snapshot of MD production run; b) Replacement of each monomer with the DFT optimized monomer, keeping the center of mass and relative orientations of each monomer the same as the MD simulation; c) Rotating the central monomer in the cluster to screen the interaction energy as a fuction of local orientation anle (ϕ), relative to the staggered configuration (ϕ=0); d) Definition of the relative orientation angle (ϕ) of middle monomer in the central trimer of selected cluster.

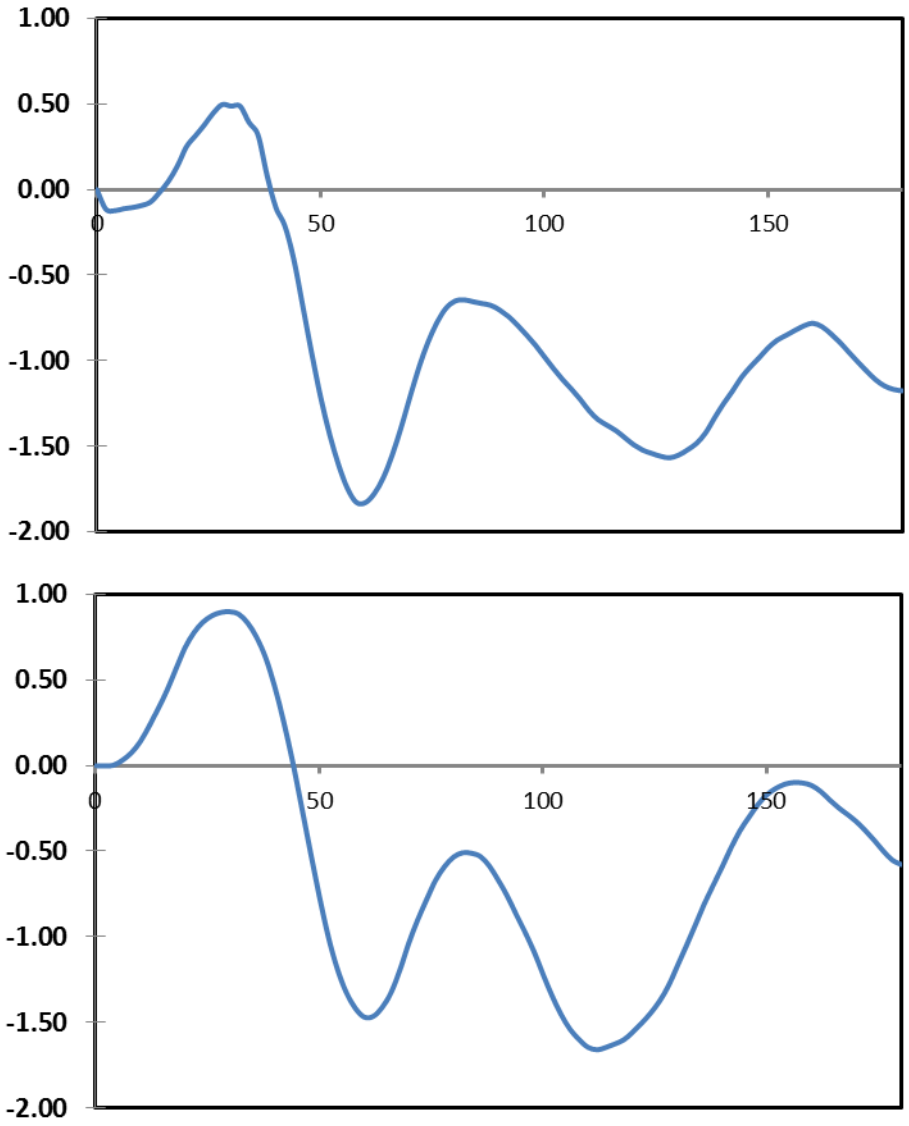


Figure B-5. Extended PES screen for sumaneneone 2 and sumanenedione 3 dimers with orientation angle from 0° to 180°. Top: sumaneneone 2; Bottom: sumanenedione 3.

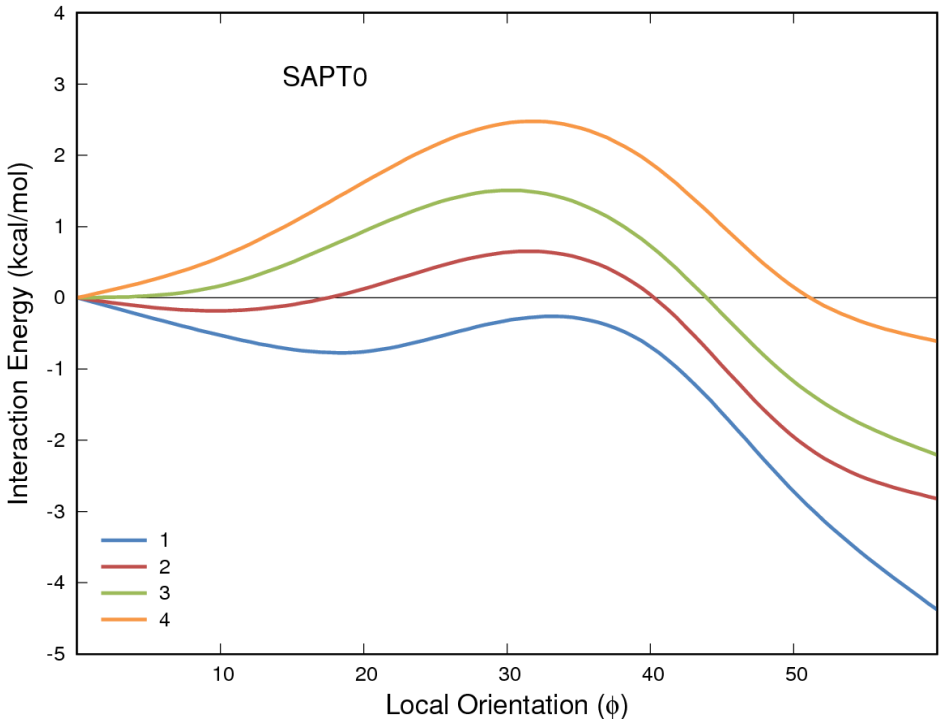


Figure B-6. jun-cc-PVDZ SAPT0 interaction energy of stacked dimers of 1 – 4 as a function of the local orientation angle (ϕ), relative to the staggered configuration (ϕ = 0) for each dimer. This plot is closely essemble the B97D/TZV(*2d,2p*) interaction energy plot in Figure 2a

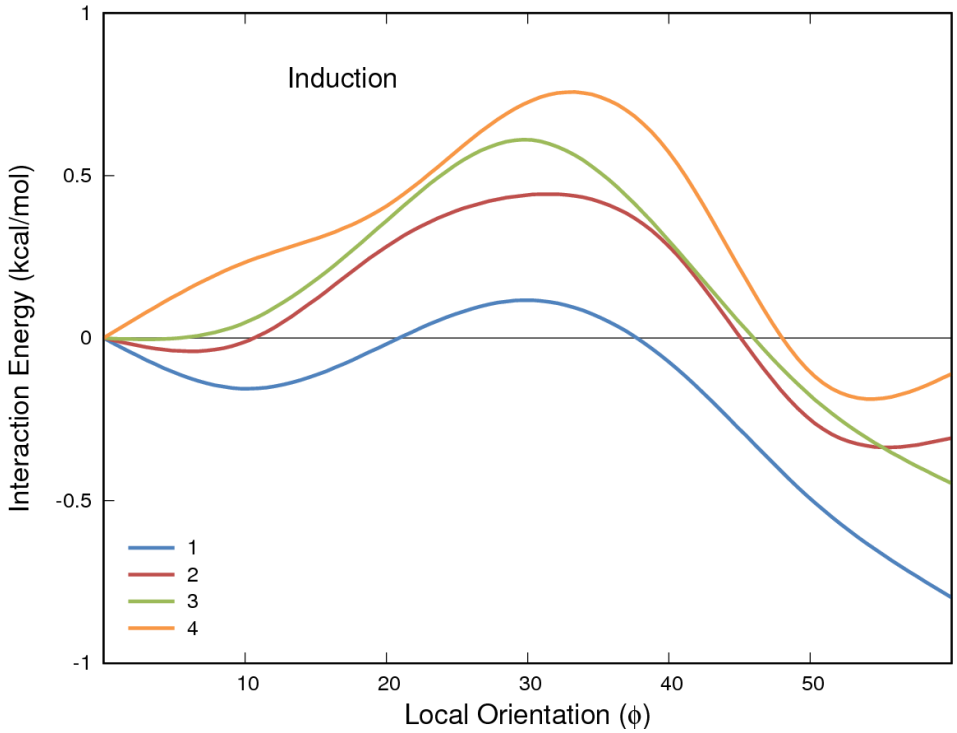


Figure B-7. jun-cc-PVDZ SAPT0 induction component of stacked dimers of 1 – 4 as a function of the local orientation angle (ϕ), relative to the staggered configuration ($\phi = 0$) for each dimer.

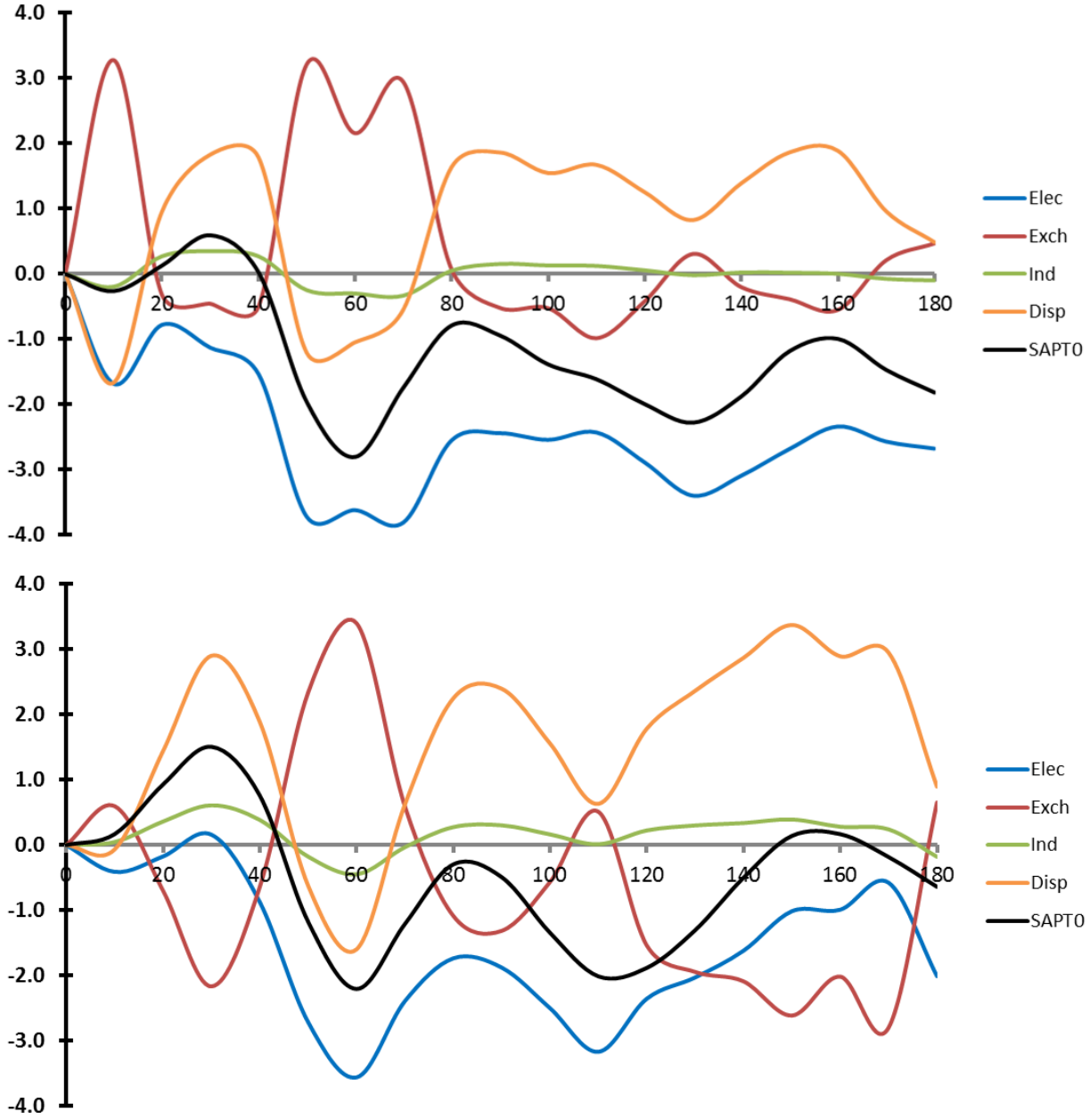


Figure B-8. SAPT interaction energies and SAPT components for stacked dimers of 2 (top) and 3 (bottom) over orientation angles from 0° to 180°.

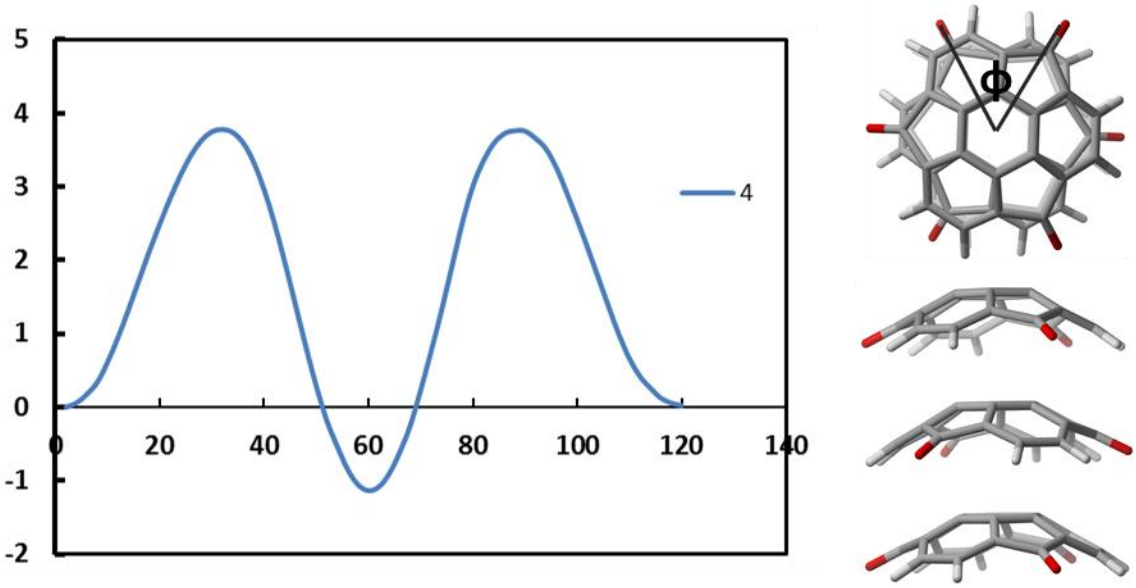


Figure B-9. B97D/TZV(*2d,2p*) interaction energy of stcked trimers of 4 as a function of the local orientation angle (ϕ), relative to the staggered configuration (ϕ=0). This result indicate that three-body stacking term can't give rise to the preference of eclipsed configuration in sumanenetrione 4 stacking structures.

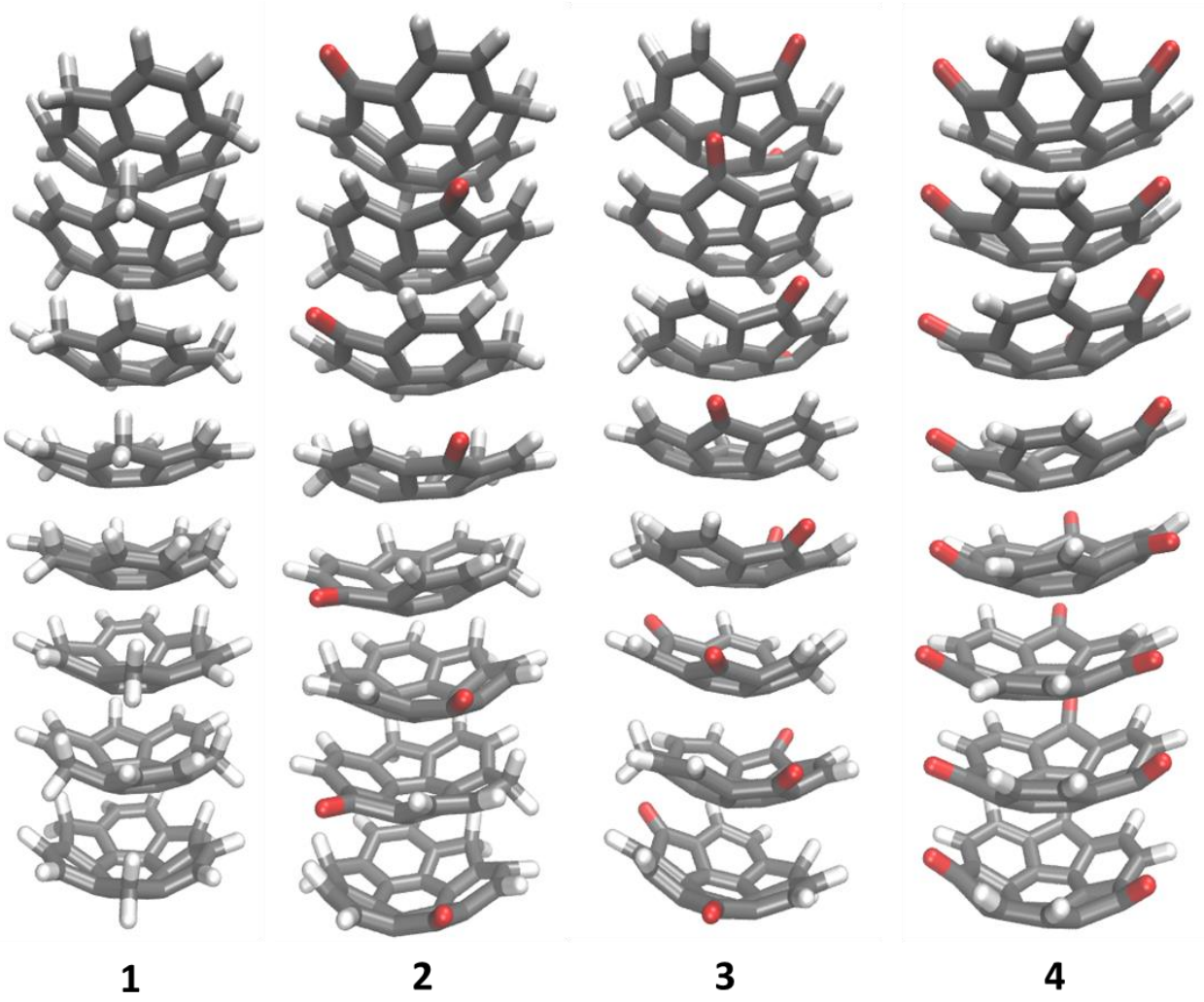


Figure B-10. Snapshots of one selected column after 100 ns production run of the MD simulation for molecules 1-4. Orientation angle can be seen from the figure as 1: ~60°, 2: 60°, 3: ~180°, 4: 0°.

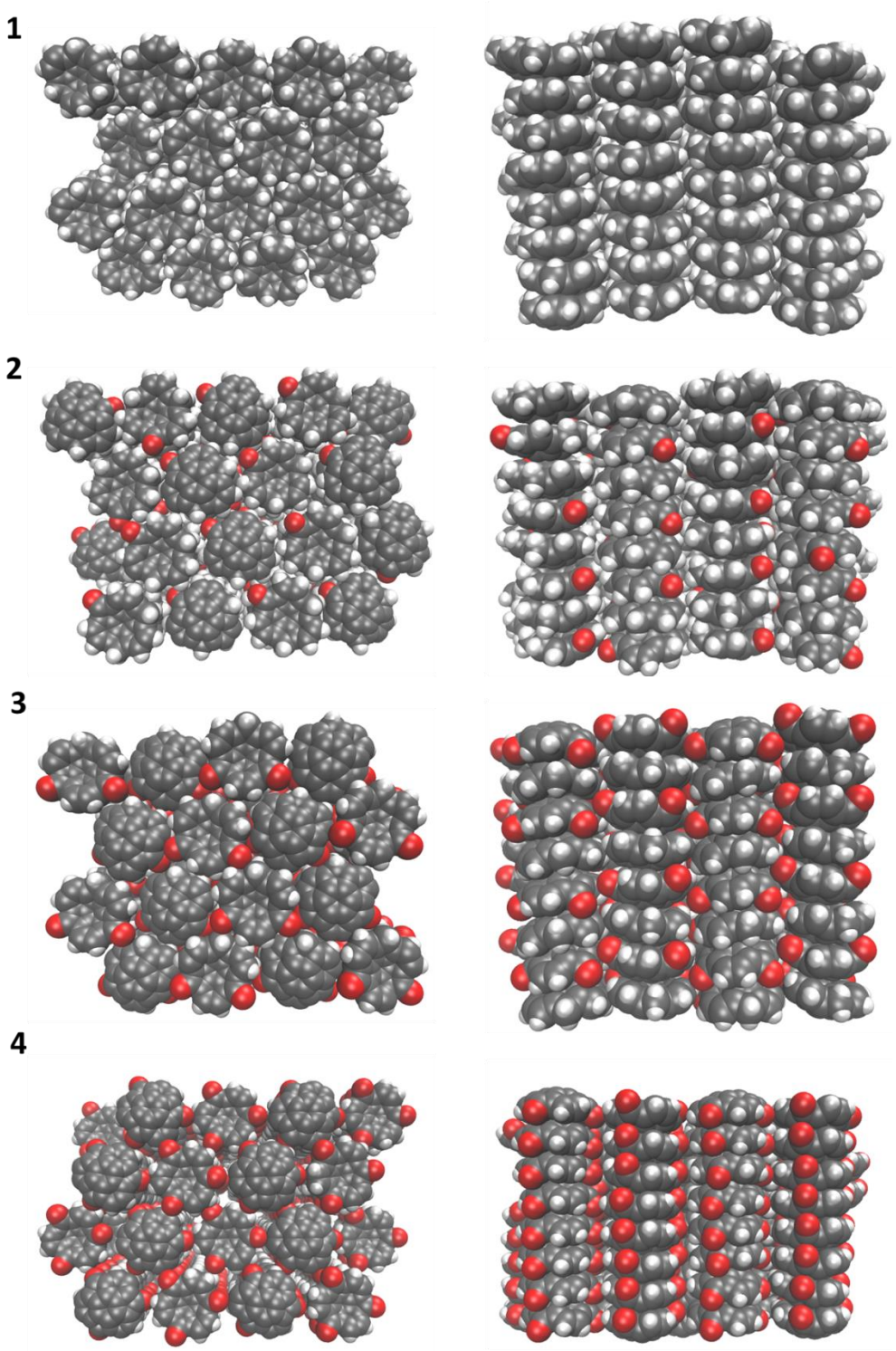


Figure B-11. Snapshots after 100 ns production run of the MD simulation for molecules 1-4.

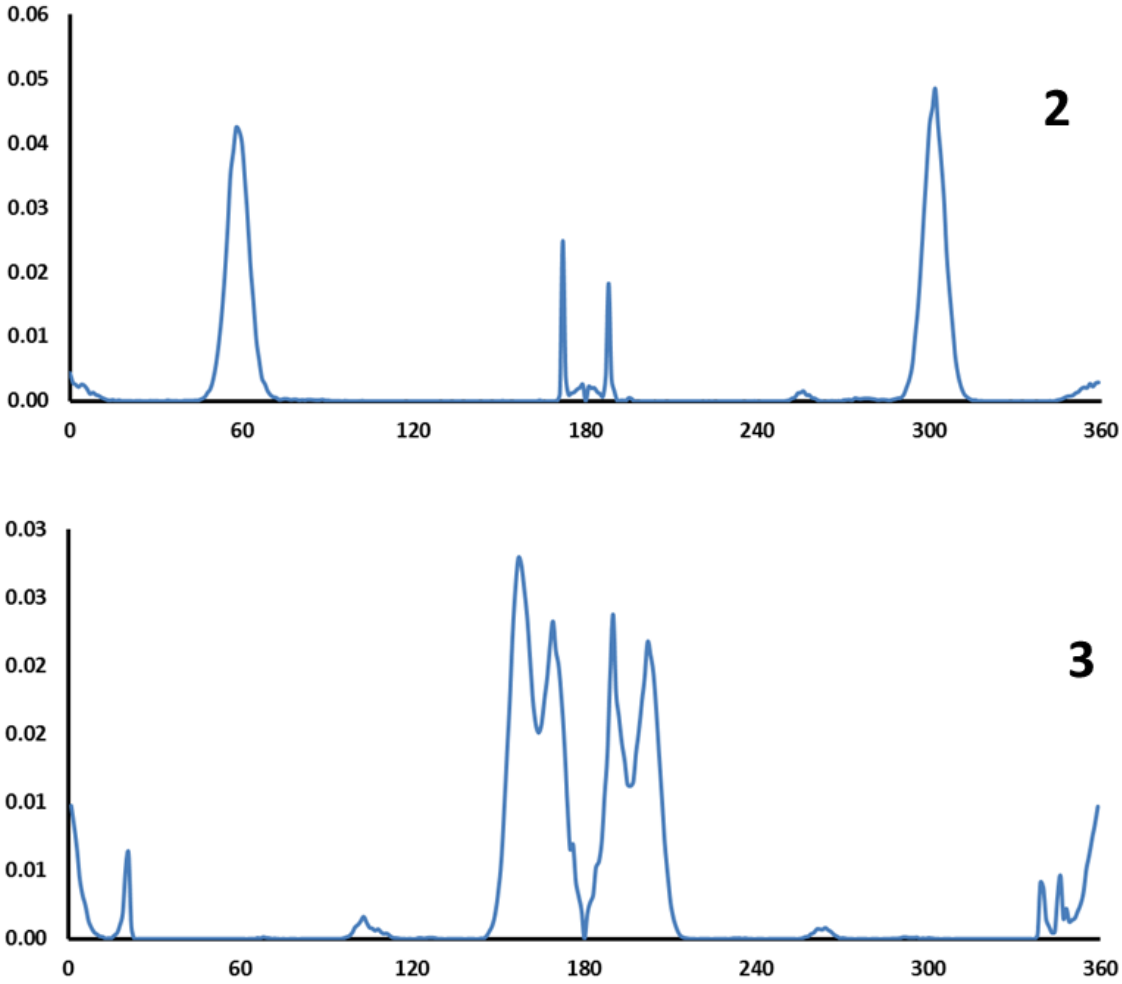


Figure B-12. Full orientation angle distribution from 0° to 360° for unsymmetrical molecules sumaneneone 2 and sumanenedione 3.

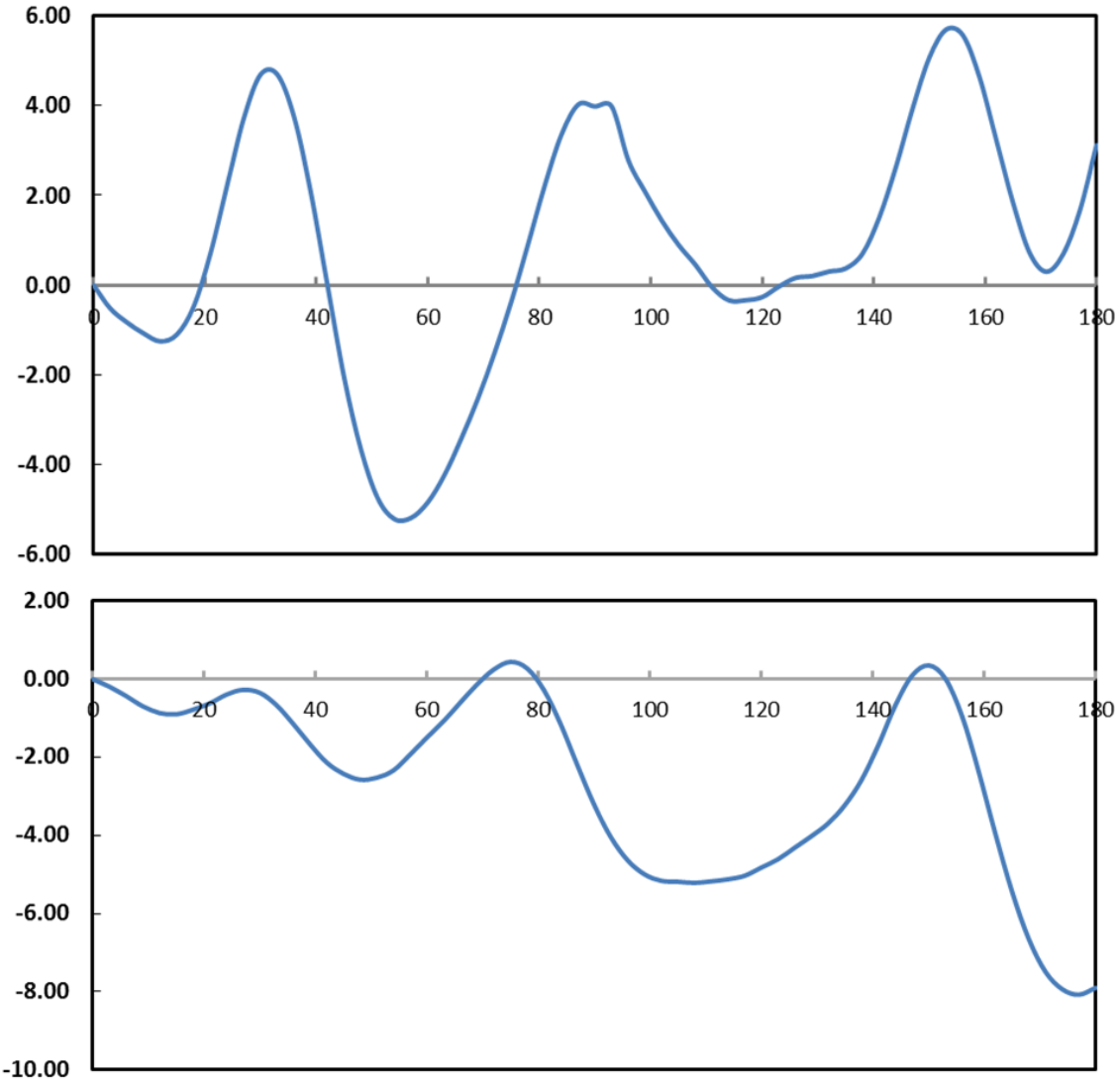


Figure B-13. Extended screened relative potential energy surface for clusters composed of 11 molecules. Top: sumaneneone 2; Bottom: sumanenedione 3.

APPENDIX C

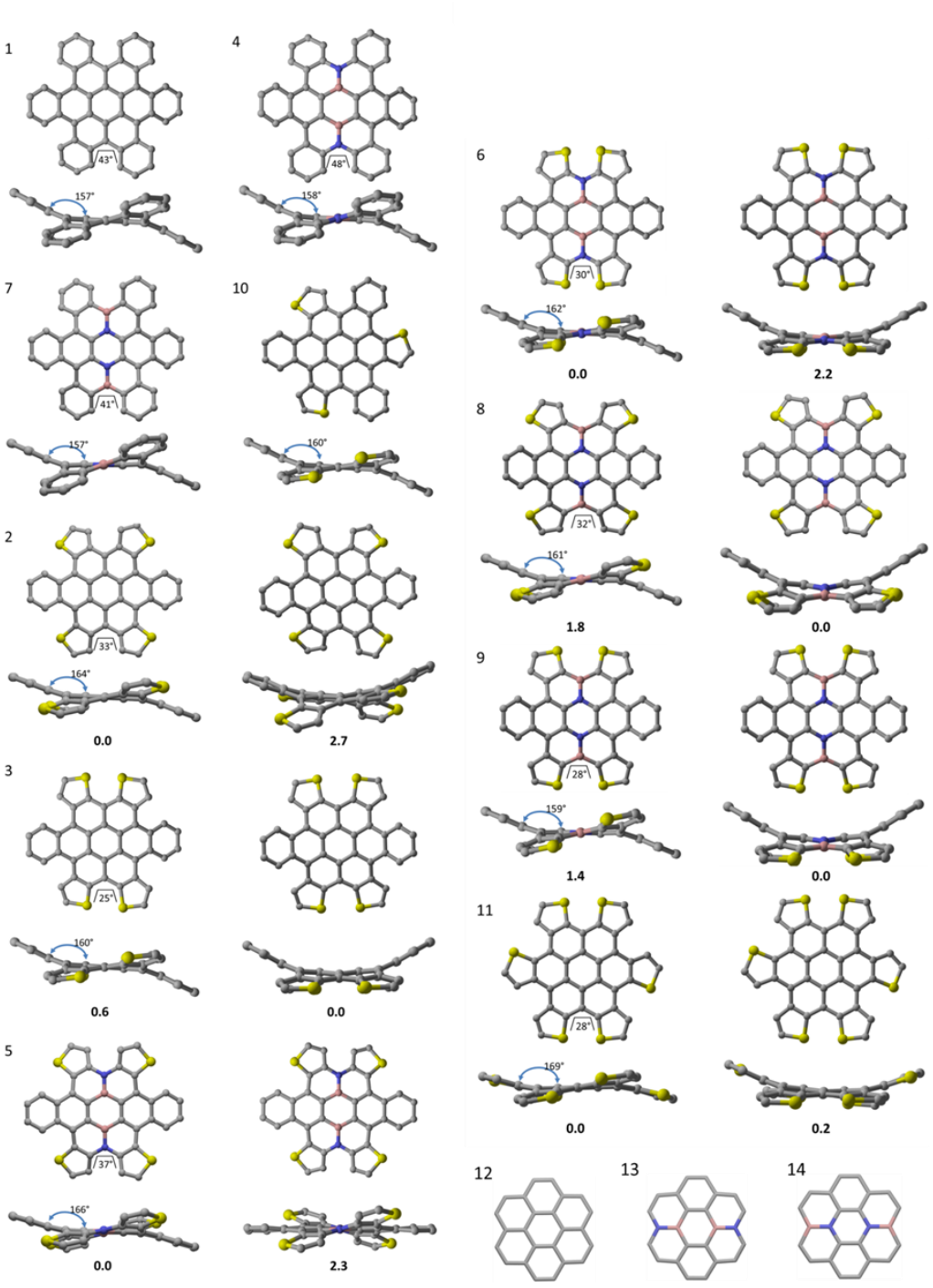


Figure C-1. Optimized structures of molecules 1 – 14.

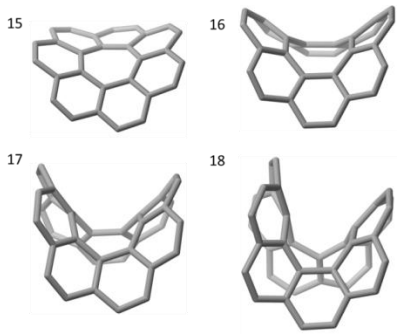


Figure C-2. Optimized structures of molecules 15 – 18.