Title
Computational Study of Enzymes and Molecular Machines

Author
Yang, Song

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Computational Study of Enzymes and Molecular Machines

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Chemistry

by

Song Yang

2018

ABSTRACT OF THE DISSERTATION

Computational Study of Enzymes and Molecular Machines

by

Song Yang

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2018

Professor Kendall N. Houk, Chair

Chapters 1-6 describe studies of enzymes and their applications in biocatalysis. Biocatalysis, the use of natural enzymes or engineered enzymes to produce molecules of interest, has been an emerging and favorable approach to supplement or replace traditional organic synthesis. Enzymes are very efficient and selective in catalysis, and enzyme catalysis is also environmentally friendly. The first portion of this dissertation reports the theoretical study of several enzymes that catalyze very different reactions, including oxidation, halogenation as well as Diels–Alder reactions. The study reveals the mechanism behind each enzyme-catalyzed reaction in terms of efficiency, substrate specificity and regio-/stereo selectivity. In addition to the understanding of each enzyme, mutations are proposed based on the mechanism of the enzyme reactions to improve the natural enzymes for better catalysis.

Chapter 1 introduces a general strategy for synthesis of macrolactones using nickel catalyzed C-C coupling and a site-selective P450 C-H oxidations carried out at Michigan by the groups of Montgomery and Sherman. Quantum mechanical (QM) computations show the intrinsic energy barriers at different hydrogen atoms at different sites in a single substrate. Molecular dynamic (MD) simulations reproduce the site selectivity and stereoselectivity in the biocatalytic oxidations with the aid of QM results. Our results suggest the linker length and its interaction with the enzyme determines the reaction yield.

Chapter 2 presents a study of the first natural Diels–Alderase (DAase) and its mechanism for catalyzing a DA reaction. QM calculations suggest the reaction goes through an ambimodal TS leading to both [4+2] and [6+4] adducts. MD simulations show that trajectories pass through the TS and go to [4+2] adduct more in the enzyme than in gas phase or water. The enzyme influences the outcome of bifurcation dramatically, mainly through hydrophobic contact. The energy barrier of the enzymatic reaction is also reproduced accurately with our newly developed method: Environment Perturbed TS Sampling (EPTSS).

Chapter 3 studies the Cope rearrangement and cyclization in hapalindole biogenesis. The X-ray crystal structure of HpiC1 is reported in this study. Mutagenesis study as well as computational computations uncover the key residues for the enzymatic reaction. QM computations show the reaction goes through an acid-catalyzed [3,3]-sigmatropic rearrangement. The switch from hapalindole to fischerindole is explained by the position change of the terminal electrophilic aromatic substitution in MD simulations.

Chapter 4 describes the mechanism of Diels–Alderase PyrI4 in pyrroindomycins biosynthesis. Density functional theory (DFT) calculations and EPTSS calculations compare the energy barriers of the reaction in the gas phase, theozyme model, and in the enzyme. Hydrogen

bonding has limited contribution to the TS stabilization in the enzyme. MD simulations show that hydrophobic interactions dominate in the catalysis of the enzyme catalyzed Diels–Alder reaction by fitting with the exo TS better than other TSs or the reactants.

Chapter 5 presents the study of a multifunctional P450 MycG and its substrate specificity. QM computations reveal the mechanism behind the MycG biopathway. MD simulations show that a hydrophobic cavity in MycG differentiates the three different substrates favors the binding of the natural substrate. Based on MD simulations, potential beneficial mutations are proposed and tested by the Sherman group at Michigan, and are proven to improve the enzyme performance in experiment.

Chapter 6 studies flavin dependent halogenase and its regioselectivity in directed evolution. The flavin dependent halogenase RebH was engineered to catalyze chlorination at different regioselective sites of tryptamine. QM calculations reveal that the intrinsic energy barriers at different sites are similar to each other. Docking and MD simulations show the different binding poses are favorable in WT and the mutants. Key mutations are identified through MD simulations and reverse mutations.

Chapters 7-9 are projects about molecular machines in solid state as well as solvent phase, studied experimentally by the Garcia-Garibay group. Ever since Prof. Feynman's famous talk "There is plenty of room at the bottom", nanotechnology has witnessed tremendous progress in the synthesis and design of molecular machines capable of mechanical movements. The early stage of molecular machine development was simply to mimic macroscopic designs. However, to succeed in building nanoscale versions of the mechanical world, more understanding is required, rather than simple mimicry. In these chapters, I have studied the dynamics of molecular rotation in dendrimetric materials, MOFs crystal and organic solvents. With computational study, the

rotational process is revealed at nanosecond time scale. New designs of molecular machine are proposed to improve their performance in different environments.

Chapter 7 studies the rotation of phenyl rings at different parts in a dendrimeric material. Molecular dynamics study reveals the different dynamics of molecular rotations at the core, branches and peripheral ends. The energy surfaces of molecular rotation are scanned using umbrella sampling, and the energy barriers are computed.

In Chapter 8, the rotational dynamics of molecular rotors in amphidynamic crystals are studied. The energy barriers of the rotation are computed by QM study to be ~0.2 kcal/mol, consistent with the ultrafast rotation in the rigid BODCA-MOF crystals. MD simulations reveal the ultrafast rotation at different temperatures and find it to be diffusion-like at high temperature.

Chapter 9 compares and studies the gearing performance of different molecular spur gears. The gearing efficiency is affected by the distance between the two rotors in molecular gears. Solvent molecules are shown to interfere with the rotation of the molecular rotors and promote slippage rather than gearing. A new design of molecular spur gear is proposed and tested by MD simulations. With a macrocyclic structure, the solvent effect is eliminated and the molecular spur gear is able to gear in solvent phase.

The dissertation of Song Yang is approved.

William M. Gelbart

Charles M. Knobler

Giovanni Zocchi

Kendall N. Houk, Committee Chair

University of California, Los Angeles

2018

In honor of my family, my love and every single person who helped me along my road

TABLE OF CONTENTS

List of Schemes

List of Figures

List of Tables

Acknowledgement

First of all, my enormous gratitude goes to Professor Kendall N. Houk, advisor of my graduate studies for the past 4 years. He took me to the group when I rediscovered my research interest in enzymes after struggling in graduate school for one and a half years. Under his guidance, I was able to learn computational chemistry very quickly and complete many projects in a productive way. He passes his enduring passion towards science to me and leads me to push back the frontier of science every day. He encourages me to get into projects with very diverse backgrounds and this has taught me to solve problems with many different perspectives. He is excellent at scientific communication both to the experts and the public, and he is always teaching me his communication skills for giving research talks and posters. In a word, he is a role model as a scientist for me, and I wish I could live up to his expectations in the future.

I am fortunate to be surrounded by a group of smart and nice colleagues in graduate school. They made every day in my graduate life enjoyable and productive by keeping inspiring me in every aspect. I have extraordinary experience with my first mentors, Dr. Gonzalo Jimenez and Dr. Jiyong Park, who taught me memorable lessons. Dr. Jimenez has left the Houk group about the time I joined, but still helped me to gain the necessary computational methods via video meetings and emails. Dr. Jiyong Park was very patient to solve all my technical problems and always sit through all of my practices on computational projects during my first year. He also taught me how to use bash language to save lots of tedious work encountered by every computational chemist.

I want to give my deep gratitude to Dr. Zhongyue Yang and Dr. Jacob Sanders, my close collaborators and friends. They have been engaging in several collaborations with me and helping me make progress by many meaningful and important discussions. They are also my gym buddies

who help spot me when I squat or bench press heavy weights. I thank them for helping me overcome barriers in both research and body building.

I am very grateful to work with and learn from a fabulous group here at UCLA, including Dr. Cyndi He, Dr. Peiyuan Yu, Dr. Marc Garcia-Borras, Dr. Jessica Grandner, Wei Li, Adam Simon, Janice Lin, Xiaofei Dong, Ieva Liepuoniute, Kersti Haatveit, and Tao Liu. They are wonderful teammates to work with, and I truly appreciate their work and contribution to my projects described herein. I have mentored several group members, Tao Liu, Ieva Liepeuniute, Kersti Haatveit, Declan Evans and Nancy Guan. I thank them for being part of my mentoring experience and letting me guide their research at the beginning.

I have many wonderful collaborators from different backgrounds, Prof. Miguel Garica-Garibay, Prof. David Sherman, Prof. John Montgomery, Prof. Jared Lewis, Prof. Michael Jung and Prof. Caius Radu. I am very thankful to all of them for being very trustful and dedicated in our collaborations. I thank Prof. Miguel Garica-Garibay for mentoring and guiding me in my first project in molecular machines. Through him and his excellent students, Dr. Xing Jiang and Marcus Jellen, I have learned to communicate with the experimental group and have fruitful results in both computation and experiments. I would also like to thank Prof. David Sherman as well as Prof. John Montgomery for having me in many of the joint projects and our awesome work on understanding enzyme reactions and enzyme engineering. I thank their students, Dr. Michael Gilbert and Dr. Matthew DeMars for working closely with me and carrying out extra experimental work based on my computational results. Thanks to Prof. Jared Lewis and his student Mary Andorfer for a very interesting project in understanding the mechanism of regioselectivity shift in directed evolution. Special thanks go to Prof. Michael Jung, Prof. Caius Radu and their joint student Ethan Rosser for having me in my very first collaboration in drug development.

I would like to show my gratitude to my committee members, Prof. William Gelbart, Prof. Giovanni Zocchi and Prof. Charles Knobler for their guidance in my early years in graduate school. It has always been enjoyable to discuss with them about teaching, research and life as a graduate student. I thank them for guiding me in discovering my research interest in biomacromolecules.

I thank my parents and my family for always supporting me through my life. They have always backed me for my decision to continue my study abroad even if it means I am thousands of miles away from them. Last but not least, I must give my gratitude to Cyndi, for giving me all of her love and support.

Chapter 1 is a modified version of the publication "Synthesis of Diverse 11- and 12-Membered Macrolactones from a Common Linear Substrate Using a Single Biocatalyst." Gilbert, M. M.; DeMars, M. D.; Yang, S.; Grandner, J. M.; Wang, S.; Wang, H.; Narayan, A. R. H.; Sherman, D. H.; Houk, K. N.; Montgomery, J. ACS Cent. Sci. 2017. This is a collaboration with the organic synthesis group and chemical biology group at University of Michigan. Dr. Michael Gilbert, Hengbin Wang and Shoulei Wang carried out/supervised the synthetic chemistry. Dr. Matthew DeMars, Dr. Michael Gilbert and Alison Narayan performed the biocatalytic oxidations and product analysis. Dr. Jessica Grandner and I performed extensive theoretical calculations on the reactions with QM and MD simulations.

Chapter 2 is a modified version of the publication "Influence of Water and Enzyme SpnF on the Dynamics and Energetics of The Ambimodal [6+4]/[4+2] Cycloaddition" Yang, Z.; Yang, S.; Yu, P.; Li, Y.; Doubleday, C.; Park, J.; Patel, A.; Jeon, B.; Russell, W. K.; Liu, H.; et al. PNAS 2018, 201719368. I carried out docking and classical molecular dynamics study of the TS enzyme complex. Dr. Zhongyue Yang, Dr. Peiyuan Yu and Dr. Yanwei Li carried out reaction dynamics study and QM/MM energy computation. Dr. Charles Doubleday and Prof. K. N. Houk designed the research.

Chapter 3 is a modified version of the publication "Structural Basis of the Cope Rearrangement and Cyclization in Hapalindole Biogenesis" Newmister, S. A.; Li, S.; Garcia-Borràs, M.; Sanders, J. N.; Yang, S.; Lowell, A. N.; Yu, F.; Smith, J. L.; Williams, R. M.; Houk, K. N.; Sherman, D. H. Nature Chemical Biology 2018, 1.

Chapter 4 describes a mechanistic study of an enzyme catalyzed Diels–Alder cycloaddition and the contribution from hydrophobic contact and hydrogen bonding to the TS stabilization. Wei Li carried out QM study of the energy pathway in the gas phase and theozyme

model. Wei Li and I also carried out classical MD simulations of TS-enzyme complex. Dr. Peiyuan Yu, Dr. Zhongyue Yang and Dr. Yanwei Li performed QM/MM calculation of energy barriers in enzyme.

Chapter 5 describes a mechanistic study of a multifunctional P450 MycG and its substrate specificity. It is a joint effort between the Houk group and Sherman group at University of Michigan. Dr. Jessica Grandner carried out QM calculation of the oxidation reactions in the gas phase. I performed MD simulations of enzyme-substrate complex, as well as metadynamics study of substrate binding into the enzyme. Dr. Matthew DeMars performed enzyme mutagenesis experiments.

Chapter 6 describes a mechanistic study of a flavin-dependent halogenase RebH and its regioselectivity in directed evolution. It is a joint effort between the Houk group and Lewis group at University of Chicago. Dr. Cyndi He and Dr. Colin Lam carried out QM calculation of the chlorination at different regioselectivity in the gas phase. I performed substrate docking into the enzyme and MD simulations of enzyme-substrate complex. Dr. Mary Andorfer performed directed evolution and enzyme reverse mutagenesis experiments.

Chapter 7 is a modified version of the publication "Crystal Fluidity Reflected by Fast Rotational Motion at the Core, Branches, and Peripheral Aromatic Groups of a Dendrimeric Molecular Rotor." Jiang, X.; O'Brien, Z. J.; Yang, S.; Lai, L. H.; Buenaflor, J.; Tan, C.; Khan, S.; Houk, K. N.; Garcia-Garibay, M. A. J. Am. Chem. Soc. 2016, 138 (13), 4650–4656.

Chapter 8 is a modified version of the publication "Ultrafast Rotation in an Amphidynamic Crystalline Metal Organic Framework" Vogelsberg, C. S.; Uribe-Romo, F. J.; Lipton, A. S.; Yang, S.; Houk, K. N.; Brown, S.; Garcia-Garibay, M. A. PNAS 2017, 201708817.

Chapter 9 describes a mechanistic study of gearing and slippage in molecular machines and design of perfect molecular spur gears working under different environments. Dr. Peiyuan Yu and Tao Liu carried out QM computations of molecular gears in the gas phase. Tao Liu and I performed classical MD simulations in the gas phase as well as organic solvents. Dr. Jiyong Park and I performed 2D metadynamics scan of rotational angles.

VITA

EDUCATION

University of California, Los Angeles Los Angeles, CA
- Ph.D. Candidate, Chemistry Jun 2018

Fudan University Shanghai, China
- B.A., Chemistry Jun 2013

RESEARCH EXPERIENCE

UCLA, Department of Chemistry and Biochemistry
Los Angeles, CA
Graduate Student with Prof. K. N. Houk Fall 2013-present

Applied molecular dynamics (MD), density functional theory (DFT) and QM/MM methods to study various systems from enzymes, drug molecules, MOFs, molecular machines to zeolite.

Enzymes
- Elucidated the mechanism of enzymatic reactions of P450 enzymes *(collaboration with Prof. David Sherman)*, halogenase *(collaboration with Prof. Jared Lewis)* and Diels-Alderases with MD simulations
- Guided the enzyme engineering of P450s for alternative substrate and broader substrate scope *(collaboration with Prof. David Sherman)*
- Developed comparative constrained MD (CCMD), a MD-based method to predict beneficial mutations for industrial enzymes *(collaboration with Codexis)*

Drug design *(collaboration with Prof. Michael Jung)*
- Identified molecular rigidity as a key factor for the binding affinity of inhibitor molecules to deoxycytidine kinase (dCK), a therapeutic target for different types of cancer

Metal-organic frameworks (MOFs) and molecular machines *(collaboration with Prof. Miguel Garcia-Garibay*)
- Revealed the dynamics of highly correlated molecular rotations in MOFs, and computed the energy barrier of rotation in solid state
- Discovered solvent effect on the gearing of molecular spur gears, and proposed a new molecular spur gear design with high gearing efficiency in solvents

Metal-doped zeolite catalyst *(collaboration with NREL)*
- Screened transition metals for metal-doped zeolite to catalyze isobutane dehydrogenation using DFT, and identified Co and Ni as promising metals for active metal-doped zeolite catalyst

PUBLICATIONS

- Jiang, X.; O'Brien, Z. J.; **Yang, S**.; Lai, L. H.; Buenaflor, J.; Tan, C.; Khan, S.; Houk, K. N.; Garcia-Garibay, M. A.: "Crystal Fluidity Reflected by Fast Rotational Motion at the Core,

Branches, and Peripheral Aromatic Groups of a Dendrimeric Molecular Rotor" *J. Am. Chem. Soc.* **2016**, *138 (13)*, 4650–4656.

- Gilbert M. M.; DeMars, M. D.; **Yang, S.**; Grandner, J. M.; Wang, S.; Wang, H.; Narayan, A. R.; Sherman, D. H.; Houk, K. N.; "Synthesis of Diverse 11- and 12-Membered Macrolactones from a Common Linear Substrate Using a Single Biocatalyst" *ACS Cent. Sci*. **2017**, 1304-1310**.**
- Vogelsberg, C. S.; Uribe-Romo, F. J.; Lipton, A.; **Yang, S.**; Houk, K. N.; Brown, S.; Garcia-Garibay, M. A.: "Ultrafast Rotations in a MOF Crystalline Material" *Proc. Natl. Acad. Sci.* **2017**, 201708817.
- Yang, Z.; **Yang, S. (co-first author)**; Yu, P.; Li, Y.; Park, J.; Patel, A.; Liu H.; Doubleday, C.; Houk, K. N.: "Influence of water and enzyme SpnF on the dynamics and energetics of the ambimodal [6+4]/[4+2] cycloaddition" Accepted by *Proc. Natl. Acad. Sci.* **2018**, 201719368.
- Newmister, S. A.; Li, S.; Garcia-Borràs, M.; Sanders, J. N.; **Yang, S.**; Lowell, A. N.; Yu, F.; Smith, J. L.; Williams, R. M.; Houk, K. N.; Sherman, D. H.: "Structural basis of the Cope rearrangement and C−C bond-forming cascade in hapalindole/fischerindole biogenesis" *Nature Chemical Biology* **2018**, 1.
- **Yang, S.**; Li, W.; Yang, Z.; Yu, P.; Houk, K. N.: "Role of Hydrogen Bonding and Steric Constraint in Enzyme Catalyzed Diels-Alder Reaction" *Manuscript in preparation.*
- **Yang, S.**; Grandner, J. M.; DeMars M. D.; Houk, K. N.; Sherman D. H.: "Molecular Dynamics Study of Substrate Selectivity by a Multi-functional P450 MycG". *Manuscript in preparation.*
- **Yang, S.**; Liu, T.; Jellen, M.; Jiang, X.; Yu, P.; Park, J.; Garcia-Garibay, M. A.; Houk, K. N.; "Molecular Dynamics Study of Gearing and Slippage Motions in Molecular Spur Gears". *Manuscript in preparation.*

FELLOWSHIPS AND AWARDS

- UCLA Seaborg Symposium Poster Award
- Sun Hung Kai Scholarship (2/95)
- National Scholarship of China (1/95)
- Scholarship Elite Liu Yong Ling (3/231)

TEACHING EXPERIENCE

UCLA, Department of Chemistry and Biochemistry
Los Angeles, CA
Teaching Assistant, Chemistry
Winter 2015-present

- Quantum Chemistry (Chem 110B)
- Statistical Mechanics (Chem 113A)

Chapter 1. Synthesis of Diverse 11- and 12-Membered Macrolactones from a Common Linear Substrate Using a Single Biocatalyst

1.1 Abstract

The diversification of late stage synthetic intermediates provides significant advantages in efficiency in comparison to conventional linear approaches. Despite these advantages, accessing varying ring scaffolds and functional group patterns from a common intermediate poses considerable challenges using existing methods. The combination of regiodivergent nickel-catalyzed C−C couplings and siteselective biocatalytic C−H oxidations using the cytochrome P450 enzyme PikC addresses this problem by enabling a single late-stage linear intermediate to be converted to macrolactones of differing ring size and with diverse patterns of oxidation. The approach is made possible by a novel strategy for siteselective biocatalytic oxidation using a single biocatalyst, with site selectivity being governed by a temporarily installed directing group. Site selectivities of C−H oxidation by this directed approach can overcome positional bias due to C−H bond strength, acidity, inductive influences, steric accessibility, or immediate proximity to the directing group, thus providing complementarity to existing approaches.

1.2 Introduction

Diversifying chemical structures at a late stage of a synthetic sequence provides an efficient way to tailor molecular properties. Effective strategies typically involve either modifying the framework of a core structural scaffold or introducing new substituents through functional group interconversions and C−H functionalization methods. Core scaffold modifications are often achieved by accessing multiple cycloaddition pathways available to a single substrate or by catalyzing a thermodynamically driven rearrangement such as transesterification, sigmatropic rearrangement, or ring expansion or contraction.[1−5] Alternatively, the introduction of new

substituents can be accomplished through the interchange of functional groups after a fully functionalized core structure is accessed. A disadvantage of this approach is that assembly of the late-stage common intermediate often requires a lengthy linear synthesis and provides structures that are closely related with an identical core framework. Site-selective C−H functionalization methods provide greater overall efficiency,[6–12] but limitations in both site selectivity and functional group compatibility (i.e., alkenes or basic amines) remain as a considerable challenge in complex environments with multiple potentially reactive centers. Biocatalytic C−H oxidations can provide impressive levels of site selectivity and functional-group compatibility, but this advantage typically requires extensive optimization of enzyme structure in order to fundamentally change the site of functionalization. Considering the array of available diversification approaches described above, a significant gap exists in synthetic strategies that involve late-stage scaffold variation combined with a versatile site-selective C−H functionalization approach. By first converting a single substrate into two or more molecular scaffolds, and then accomplishing site-selective diversification of each of the obtained structural motifs through C−H functionalization, the rapid assembly of structural analogues differing in both the core scaffold and peripheral functionality would become possible (Figure 1). Such an approach offers the simplicity of accessing a linear substrate with minimal functionality as the point of divergence and would provide a powerful strategy for the rapid assembly of molecular diversity and complexity. In this report, we describe proof-of-concept experiments toward this objective with the demonstration of a new strategy that enables regiodivergent cyclization paired with site-selective biocatalytic oxidations. This approach provides access to a collection of 11- and 12-membered macrocycles with varying oxidation patterns.

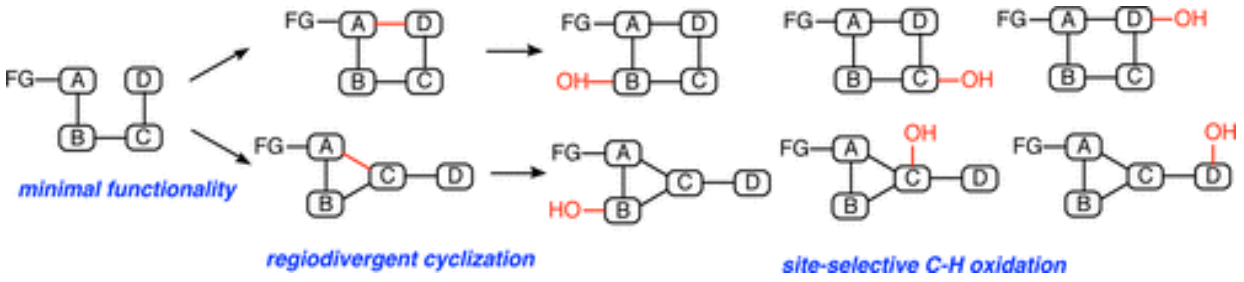


Figure 1.1. Conceptual framework for late-stage diversification. Regiodivergent scaffold assembly followed by site-selective C–H oxidation presents a versatile strategy for accessing diverse products from a single late-stage intermediate.

1.3 Methods

General Procedure for Reductive Macrocyclization

In a glovebox, an oven-dried flask equipped with stir bar was charged with 33 mol % of $Ni(COD)^2$ and 30 mol % of IPrCl. The flask was capped with a septum, removed from the glovebox, and attached to a nitrogen line. The contents of the flask were suspended in toluene (6.3 mL) and stirred for 20 min. The contents were diluted with an additional 31.4 mL of toluene and heated to 60 °C in an oil bath for 20 min. Triisopropylsilane (5 equiv) was then added in a single portion. Ynal (1 equiv) was then added by syringe drive over 1 h as a 4.15 mL solution in toluene. The reaction mixture was stirred overnight at 60 °C. The reaction mixture was cooled to room temperature, and the volatiles were removed under vacuum. The crude residue was filtered through a plug of silica with EtOAc/hexanes (1:1) and then purified by column chromatography.

General Procedure for Biocatalytic Oxidation

To an Erlenmeyer flask containing reaction buffer (50 mM NaH_2PO^4, 1 mM EDTA, 0.2 mM DTT, 10% (v/v) glycerol, pH 7.3) were added the following components sequentially: substrate (20 mM stock in DMSO, 1 mM final concentration), glucose-6-phosphate (100 mM stock in reaction buffer, 5 mM final concentration), glucose-6-phosphate dehydrogenase (100 U/mL stock in water, 1 U/mL final concentration), PikC-RhFRED (varied stock concentrations, 5 μM

final concentration), and $NADP^+$ (20 mM stock in reaction buffer, 1 mM final concentration). The reaction mixture was capped with a milk filter and incubated at 30 °C overnight (14–16 h) with gentle shaking (100 rpm). Reactions were typically conducted on ~40–60 mg of each substrate (~90–130 mL total reaction volume) and performed in 500 mL Erlenmeyer flasks. After overnight incubation, the reaction was quenched by addition of acetone (2× total reaction volume) and incubated at 4 °C for 2 h.

QM calculations

All QM calculations were performed with Gaussian09. Conformational searches were performed as described in the main text. Geometry optimizations and frequencies calculations were performed at the B3LYP level with LANL2DZ for iron and 6-31G(d) for all other atoms. Transition structures contained one negative frequency. Enthalpies and free energies were computed at 1atm and 298.15K. A correction to the entropy was applied in accordance with the work of Truhlar et al. Single point energy computations were performed at the B3LYP-D3(BJ) level with LANL2DZ for iron and 6-311+G(d,p) and CPCM for water.

MD simulations

The haem iron(IV)-oxo complex involved in the cytochrome-catalyzed oxidative hydroxylation cycle (compound I) was used to model the active form of the PikC cofactor. Simulations were performed using the GPU code (pmemd) of the Amber 12 package. The Amber-compatible parameters developed by Cheatham et al. were used for compound I and its axial Cys ligand. Parameters for macrolactone substrates were generated within the antechamber module using the general AMBER force field (gaff), with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model. The charges were calculated according to the Merz-Singh-Kollman scheme using the Gaussian 09 package. Each protein was immersed in a

pre-equilibrated truncated cuboid box with a 10 Å buffer of TIP3P water molecules using the tleap module, resulting in the addition of around 11370 solvent molecules. The systems were neutralized by addition of explicit counter ions (Na^+ and Cl^-). All subsequent calculations were done using the widely tested Sony Brook modification of the Amber 99 force field (ff99sb). The substrate and enzyme were optimized for total 1000000 steps, with 750000 steepest descent steps and 250000 conjugate gradient steps. The systems were gently heated using six 50 ps steps, incrementing the temperature by 50 K for each step (0-300 K) under constant-volume and periodic-boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between hydrogen atoms were kept fixed. Long-range electrostatic effects were modelled using the particle-Mesh-Ewald method. An 8 Å cutoff was applied to Lennard-Jones and electrostatic interactions. Harmonic restraints of 30 kcal/(mol·$Å^2$) were applied to the solute and the Andersen equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated for 2 ns with a 2 fs time step at a constant volume. Production trajectories were then run for an additional 500 ns under the same simulation conditions.

1.4 Results and Discussion

A key goal of this study was to develop a versatile oxidation method that utilizes a single biocatalyst to provide multiple oxidation patterns selectively, enabled by the design of directing groups temporarily installed on the substrate. To address this objective, preparation of test substrates by a nickel-catalyzed reductive macrocyclization process allows access to either regiochemical outcome (i.e., endo- or exocyclization) in macrolactone assembly by tailoring the ligand structure and reaction conditions (Figure 2A).13,14 This regiodivergent catalytic process provides an ideal approach for accessing different ring sizes of macrolactone substrates that

possess multiple unactivated methylene (CH_2) groups for exploration of new strategies for site-selective C−H oxidation. While impressive strides have been made in strategies that enable siteselective oxidations in multi-ring structures with well-defined conformations,[15–18] selective access to multiple patterns of oxidation in macrocyclic compounds represents a challenge unmet by previous approaches. In order to address this limitation, the straightforward modular assembly of an enantio pure ynal precursor 1 was followed by regiodivergent and highly diastereoselective reductive macrocyclization to afford macrocycles 2 and 3. This approach, governed largely by steric properties of the N-heterocyclic carbene (NHC) ligand, enables access to a large array of linear and macrocyclic structures through predictable reversals of regiochemistry in the C−C bond-forming step.

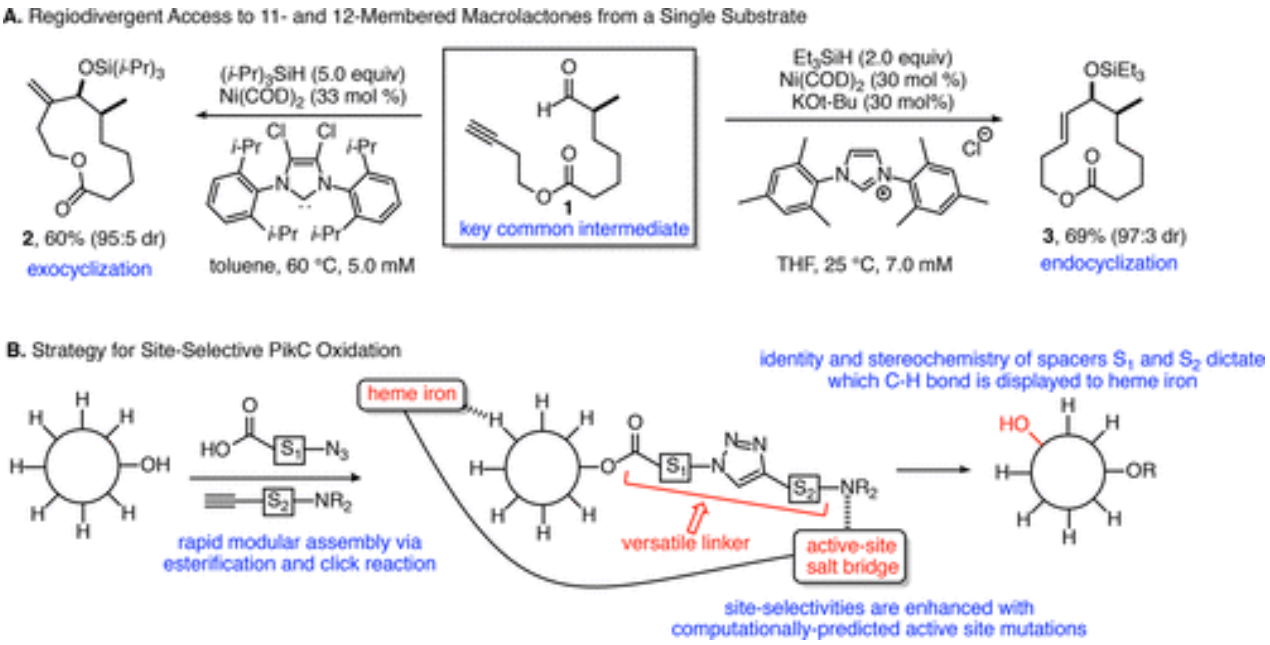


Figure 1.2. Synergy of small molecule catalysis and biocatalysis in late-stage diversification. (A) Regiodivergent reductive macrocyclization enabling access to 11- and 12-membered macrolactones. (B) Modular strategy for enabling site-selective biocatalytic oxidation with the potential to alter site of oxidation based on linker design.

Macrocyclic substrates 2 and 3 provide a challenging context for developing site-selective C−H oxidations on different scaffolds that possess a multitude of similarly reactive C−H bonds.

To address this challenge, we examined the utility of an engineered fusion protein of the bacterial cytochrome P450 enzyme PikC. In prior work, engineered mutants of the PikC enzyme displayed excellent site selectivity in oxidations of amine-containing substrates closely related to the endogenous macrolide substrate YC-1719 or of smaller, structurally compact substrates such as simple terpenes (e.g., menthol).[20] This strategy typically provided either a single major product or the unselective generation of numerous products. Therefore, a general strategy for changing the site of oxidation while preserving high selectivity represents a major unmet need from the prior approaches. The conformational flexibility of macrocycles such as 2 and 3 that possess six or more methylene groups poses a tremendous challenge in siteselective oxidation that has not yet been efficiently addressed by chemical or biological catalysis. With such subtle differences in the reactivity of similar methylene groups, the use of directing groups presents the best opportunity for achieving tunable site selectivity. While recent developments in enzyme engineering have demonstrated remarkable advances in organic synthesis,[21–24] the use of simple directing groups to achieve site selectivity in enzymatic transformations is greatly underutilized. Structural variations in rigid directing groups have recently demonstrated enormous promise in remote C−H functionalizations using small molecule catalysts.[25,26] Despite pioneering early work from Breslow in designing biomimetic site-selective functionalizations based on template design,[27,28] this approach has not previously been developed as a strategy to enable multiple selective outcomes in an enzyme-catalyzed process.

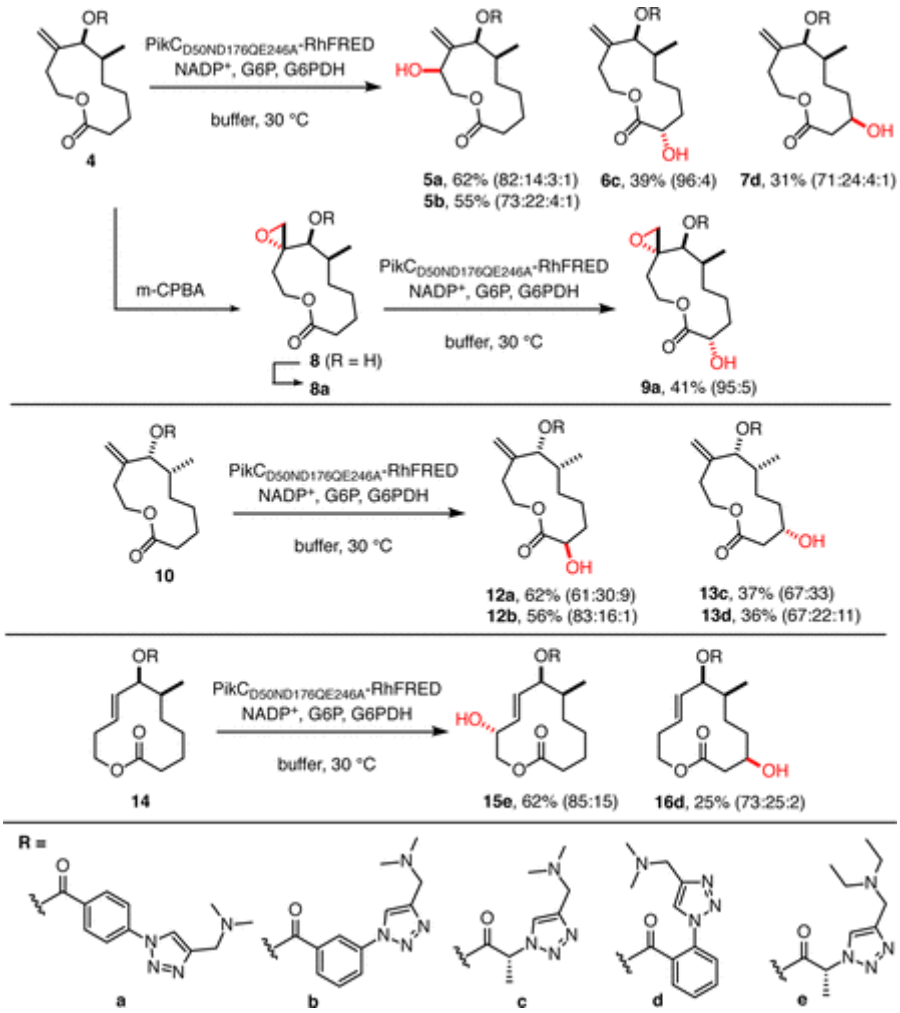
Previous studies of PikC-catalyzed oxidations illustrated that substrates possessing numerous contiguous methylene groups led to unselective oxidation when substrate engineering approaches were attempted.[29] Thus, although the presence of the aminosugar desosamine is essential for binding and selectivity of the PikC enzyme toward its native macrolide substrates, at

least seven oxidized products were generated when using a desosaminylated, but unfunctionalized, 12membered macrocyclic substrate. While the use of simple aromatic spacers avoids the cumbersome installation of aminosugars,[19] we recognized the need for a more versatile strategy that enables an anchoring basic amine to be covalently attached through a linker. The method should provide a rapid and high-throughput synthesis, facile length and shape variations of the linker component, and efficient chemoselective removal of the linker following C−H oxidation. These criteria can all be addressed by the modular connection of hydroxylcontaining substrates, amino acid-derived azido acids, and acetylenic amines (Figure 2B). Through the appropriate choice of spacer elements S1 and S2 and access to regiochemistry reversal in azide−alkyne click cycloadditions,[30,31] a large number of linker structures are readily available. Notably, click reactions are typically utilized as a simple means to connect two structures without regard to the precise structural features of the linking functionality. However, in our approach, the precise shape and geometry of the unit assembled by the azide−alkyne cycloaddition is an integral feature required for success of the strategy.

Derivatives of 11-membered substrate 2 were first examined in the click anchor strategy for site-selective biocatalytic oxidation (Table 1). The engineered PikC fusion protein PikCD50ND176QE246A-RhFRED, which was designed by molecular dynamics-guided analysis20 utilizing crystallographic insights from the endogenous substrate YC-17,[32] was used in our studies. The active site residues that were predicted to lead to unproductive substrate binding orientations were mutated in order to promote formation of the closed, active form of the enzyme. A collection of 13 triazole anchors was then assembled from seven azido acids and two amines (full listing in Table S1). Triazoles examined include those derived from azido acids possessing and lacking chirality, ortho, meta, and para-spaced benzene spacers, and both 1,4- and 1,5-triazole

regioisomers. Following installation of the triazole anchor to substrate 2, analytical scale biocatalytic oxidations on derivatives of 4 were performed, and promising anchors were selected from LCMS analysis based on percent conversion and selectivity for a major product. Representative cases were then conducted on a 30−60 mg preparative scale, products were isolated by preparative HPLC, and the structure and stereochemistry of major products were elucidated by NMR. Isolated monohydroxylated products were then used as internal standards in the analysis of additional analytical scale experiments in order to identify anchors that enable synthesis of a different hydroxylated product.

Table 1.1. Scope of Biocatalytic Macrocycle Oxidation



Using this approach, three different monohydroxylated products (5, 6, and 7) were obtained from substrate 4. Notably, triazoles possessing para- ormeta-substituted benzene spacers a or b enable oxidation of allylic protons proximal to the point of anchor connection to provide product 5. Alternatively, selection of anchors that possess a shorter linker motif enables oxidation of the distal region of the substrate. For example, using alanine-derived anchor c leads to the production of product 6c with oxidation α- to the carbonyl. In contrast, anchor d with an ortho-substituted benzene spacer leads to product 7d with oxidation β to the carbonyl. No evidence for epoxidation or amine oxidation was observed in PikC-catalyzed oxidations, demonstrating desirable chemoselectivity features of C−H oxidation. Epoxide 8 is cleanly obtained (when R = H) either in reactions with m-CPBA as the electrophilic oxidant or by using the White C−H oxidation catalyst,18 demonstrating orthogonal chemoselectivity of the PikC oxidation method. By installing anchor a, biocatalytic oxidation of 8a then cleanly affords epoxy alcohol 9a. The site-selective oxidation to provide 5, 6, 7, or 9 demonstrates that this method is useful for oxidizing C−H bonds that are either proximal or distal to the directing group. No conversion was noted when the dimethylamino group of 4a was replaced by an isopropyl or imidazole functionality, thus validating the directing capability of the dimethylamino group. The versatile directing capability of the triazole linkers paired with PikC catalysis thus does not directly correlate with C−H bond strength, acidity, inductive influences, steric accessibility, or immediate proximity to the directing group. Instead, the enzyme active site conformation matched with the structure, stereochemistry, and conformation of both the triazole anchor and the substrate overrides these influences.

To probe the interrelationship of the stereochemical features of the active site residues, triazole anchor, and macrocyclic substrate, the enantiomeric form of the macrocycle (substrate 10) was examined with the same biocatalyst and triazole anchors a−d. By using the meta- and para-

substituted anchors a and b that enabled oxidation of the proximal allylic methylene in substrate 4, these same anchors instead led to oxidation of distal protons α to the carbonyl to provide products 12a and 12b. Use of an alanine-derived anchor c or the ortho-substituted anchor d afforded 13c and 13d as the major product via oxidation β to the carbonyl. These results demonstrate that optimal anchor structure will vary from substrate to substrate, even within an enantiomeric series, thus adding appeal to the simple and potentially high throughput access to linkers in the approach described.

In taking advantage of the regiodivergent access to macrocycles described, the 12-membered structure derived from endocyclization was also examined in biocatalytic oxidations. In this case, two different oxidized products were obtained from substrate 14. The alanine anchor e provided access to allylic oxidation product 15e, and a diethylamino anchor was employed in this case to minimize Ndemethylation, which was observed as a minor byproduct in some instances. Alternatively, using the ortho-linked benzene spacer d, major product 16d was obtained via oxidation β to the carbonyl. These results collectively illustrate that a variety of monohydroxylated compounds of macrocycles varying in ring size may be efficiently obtained starting from a single, easily accessible ynal substrate.

As described above, the structure, stereochemistry, and conformation of the substrate, linker, and enzyme active site all play important roles in determining site selectivity of the C−H oxidations. In order to understand these influences and to add a predictive component to the strategy, a combined density functional theory (DFT) and molecular dynamics (MD) computational study was undertaken. DFT was used to compute the intrinsic reaction barriers to C−H abstraction at C3 allylic and C10 alpha positions of 4. The R group shown in 4 was truncated to a methyl, and conformations of the macrocycle were explored with molecular mechanics.33

These conformers were then optimized, and single point energy calculations were performed with DFT. The lowest energy conformer was identified along with a second low energy conformer only 0.9 kcal/mol higher in energy. All other conformers were >2 kcal/mol higher in energy than conformer 1. The barriers to C−H abstraction at C3 and C10 were computed for these two low energy conformers and are shown in Figure 3A. The barriers to abstraction of the hydrogens shown in purple and green are the lowest for each site of both conformers, and correspond to the hydrogens that are abstracted to form 5 and 6. Figures 3B and 3C show the transition structures for abstraction of the C3 purple (equatorial) and C3 yellow (axial) hydrogens, respectively, from conformer 1. It is clear from close inspection of the transition structures that the abstraction of the equatorial hydrogen (purple) benefits from developing conjugation with the neighboring exocyclic alkene. The axial hydrogen (yellow) is perpendicular to the π-system and cannot benefit from the allylic stabilization. A similar conjugation effect is seen with the C−H abstractions at C10. The green hydrogen is in conjugation with the carbonyl group (Figure 3D), but the ring must distort for the blue hydrogen to benefit from the same stabilization. The substrate prefers a dihedral of 16.1° between the exocyclic alkene and the carbonyl. In Figure 3E, this dihedral (shown with blue atoms) expanded to 55.6° in the transition structure for the abstraction of the blue hydrogen.

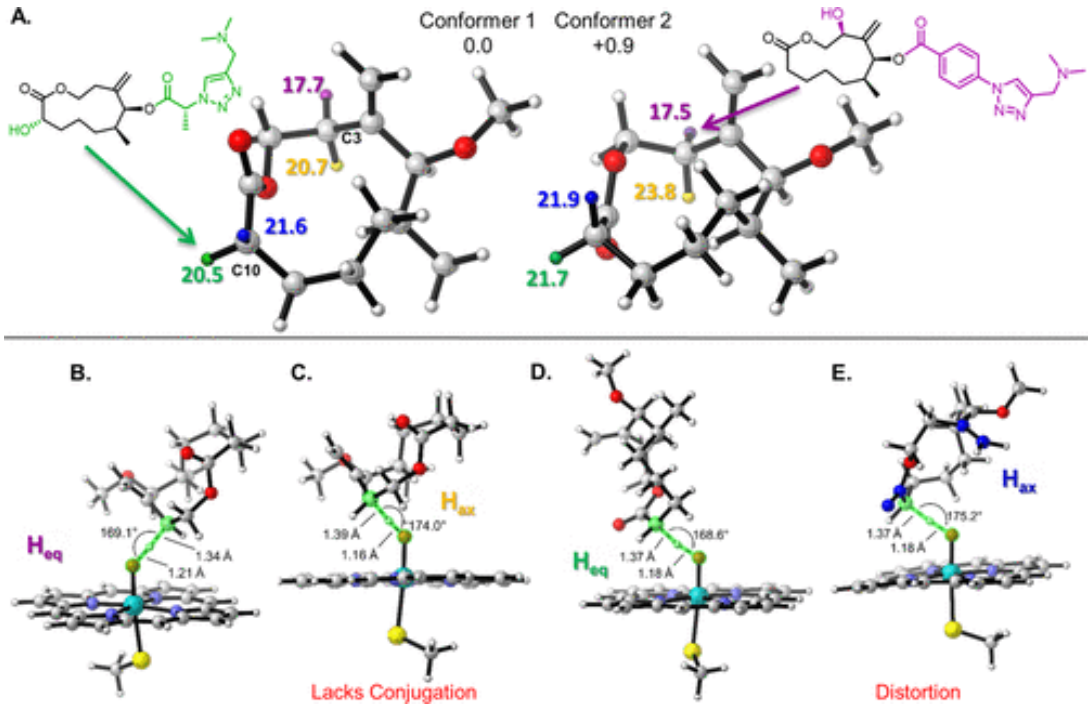


Figure 1.3. (A) Lowest energy conformers of a model of structure **4**, with DFT barriers (kcal/mol) to C–H abstraction at C3 and C10. (B) Transition structure of C3 (purple) hydrogen abstraction. (C) Transition structure of C3 (yellow) hydrogen. (D) Transition structure of C10 (green) hydrogen abstraction. (E) Transition structure of C10 (blue) hydrogen abstraction.

Conformer 1 was used for molecular dynamics simulations. Linker a and linker c were each attached to the macrocycle and docked into the active site of the PikCD50N/D176Q/E246A, and 500 ns trajectories were run on each of these complexes using Amber. Over the course of the simulations, both conformer 1 and conformer 2 are observed. The salt bridges between the terminal amines of the substrates and residues E84 and E95 are critical for substrate binding and reactions as studied in previous work by our groups.20 The same salt bridges are also observed in MD simulations of 4a and 4c. Though the salt bridges are similar for both linkers, different linkers lead to different substrate binding modes and oxidation patterns. The goal of our molecular dynamics studies here is to determine how the length, flexibility, and stereochemistry of these linkers allow for the different oxidation patterns. As shown in the snapshot of Figures 4A and 4B, linker a places C3 closest to the active iron−oxo. C2 is also placed close to the iron−oxo (3.7 Å averaged over the

500 ns simulation), but the computed barriers to abstraction at C2 are +5 kcal/mol higher in energy than at C3 (Figure S10). Alternatively, linker c places C10 closest to the iron−oxo species. Figures 4C and 4E show the average distance of each hydrogen to the iron−oxo oxygen over the course of the 500 ns simulations. Figures 4D and 4F are plots of the Hsubstrate−Oiron distances vs C−H−O angles over the course of the trajectories with linkers a and c respectively. Figures 4C and 4D reveal that the purple hydrogen is closest to the DFT-computed transition state geometry. Collectively these data show that the purple, equatorial hydrogen of C3 is the most accessible to the iron−oxo in the enzyme and intrinsically more reactive. Figures 4E and 4F show that the blue hydrogen is closest to the DFT-computed transition state geometry, but abstraction of this hydrogen is disfavored by up to 1.1 kcal/mol. While linker c places C10 closest to the iron− oxo, the intrinsic reactivity of the equatorial hydrogen (green) overrides the proximity of the axial hydrogen (blue). MD simulations of 4d and 14e were also able to reproduce the stereo preference of their corresponding major products, 7d and 15e respectively (Figures S11−S13). The combined DFT and MD study provides an explanation for the observed regiochemistry and stereoselectivity of these PikC-catalyzed hydroxylations.

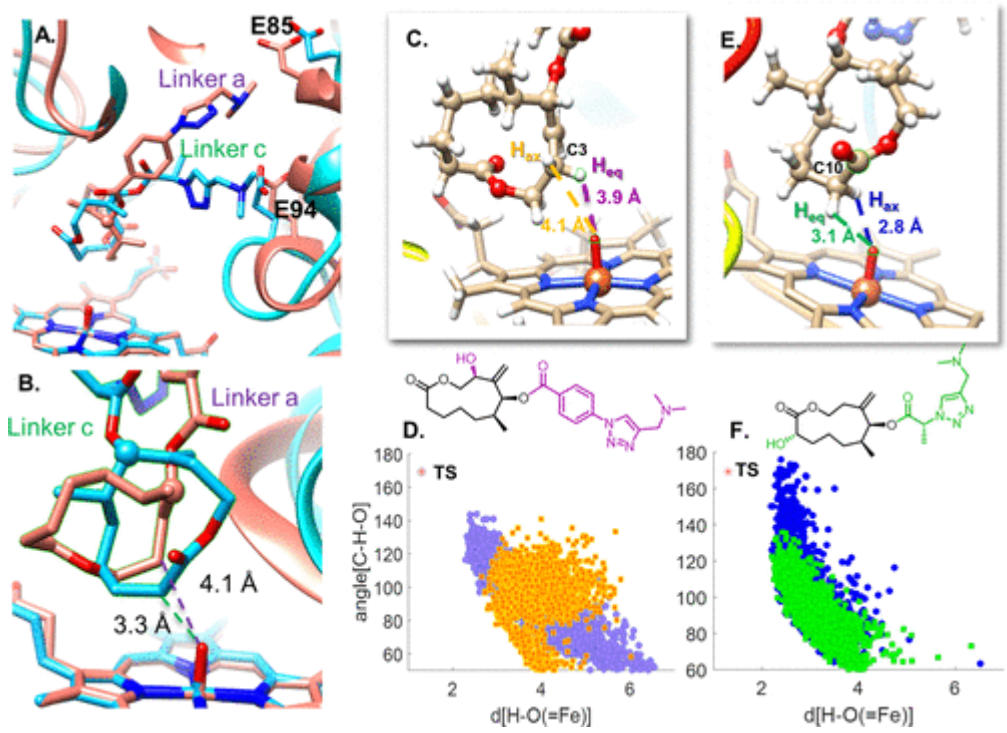


Figure 1.4. A) Snapshot of MD trajectory of **4** with linker **a** overlaid with a snapshot of **4** with linker **c**. (B) Closeup of Figure 3A snapshot with average C–O_{Fe} distances shown. (C) Snapshot of **4** with linker **a** with average H–O distances shown. (D) Plot of hydrogen (of substrate C3) to oxygen (of iron–oxo) distances vs C–H–O angles throughout the MD trajectory, with transition state (TS) geometry shown in red. (E) Snapshot of **4** with linker **c** with average H–O distances shown. (F) Plot of hydrogen (of substrate C10) to oxygen (of iron–oxo) distances and C–H–O angles throughout the MD trajectory, with TS geometry shown in red.

1.5 Conclusions

In summary, this study describes a versatile strategy for the rapid generation of a collection of macrocyclic compounds that differ in ring size and oxidation pattern. Starting from a single linear substrate, catalyst control in a regiodivergent macrocyclization is paired with site-selective C−H oxidations enabled by the synergy of a computationally designed engineered cytochrome P450 catalyst and a tailored amine-containing directing group. This new approach for controlling selectivities of biocatalytic oxidations enables a single P450 biocatalyst to be utilized for accessing multiple oxidation patterns on flexible macrocyclic motifs containing an array of oxidizable bonds. Computational analysis uncovers a complex synergy of macrocycle conformation, inherent reactivity of various C−H bonds, linker structure, and substrate-active site interactions in

controlling site selectivity of C−H oxidation. This work provides a general strategy and new opportunities for diversification through late-stage functionalization of a broad array of substrate classes.

1.6 References

1. Burke, M. D.; Schreiber, S. L. A planning strategy for diversity-oriented synthesis Angew. Chem., Int. Ed. 2004, 43, 46– 58 DOI: 10.1002/anie.200300626

2. Bos, P. H.; Antalek, M. T.; Porco, J. A.; Stephenson, C. R. J. Tandem dienone photorearrangement-cycloaddition for the rapid generation of molecular complexity J. Am. Chem. Soc. 2013, 135, 17978– 17982 DOI: 10.1021/ja409992m

3. Medeiros, M. R.; Narayan, R. S.; McDougal, N. T.; Schaus, S. E.; Porco, J. A. Skeletal diversity via cationic rearrangements of substituted dihydropyrans Org. Lett. 2010, 12, 3222– 3225 DOI: 10.1021/ol101144k

4. Nicolaou, K. C.; Petasis, N. A.; Zipkin, R. E. The endiandric acid cascade. Electrocyclizations in organic synthesis. 4. Biomimetic approach to endiandric acids A-G. Total synthesis and thermal studies J. Am. Chem. Soc. 1982, 104, 5560– 5562 DOI: 10.1021/ja00384a080

5. Zacuto, M. J.; Leighton, J. L. Divergent synthesis of complex polyketide-like macrolides from a simple polyol fragment Org. Lett. 2005, 7, 5525– 5527 DOI: 10.1021/ol052373g

6. Gutekunst, W. R.; Baran, P. S. C-H functionalization logic in total synthesis Chem. Soc. Rev. 2011, 40, 1976– 1991 DOI: 10.1039/c0cs00182a ,

7. Newhouse, T.; Baran, P. S. If C-H bonds could talk: Selective C-H bond oxidation Angew. Chem., Int. Ed. 2011, 50, 3362– 3374 DOI: 10.1002/anie.201006368 ,

8. Topczewski, J. J.; Cabrera, P. J.; Saper, N. I.; Sanford, M. S. Palladium-catalysed transannular C-H functionalization of alicyclic amines Nature 2016, 531, 220– 224 DOI: 10.1038/nature16957 , ,

9. Wencel-Delord, J.; Glorius, F. C-H bond activation enables the rapid construction and late-stage diversification of functional molecules Nat. Chem. 2013, 5, 369– 375 DOI: 10.1038/nchem.1607

10. Cernak, T.; Dykstra, K. D.; Tyagarajan, S.; Vachal, P.; Krska, S. W. The medicinal chemist's toolbox for late stage functionalization of drug-like molecules Chem. Soc. Rev. 2016, 45, 546–576 DOI: 10.1039/C5CS00628G

11. Durak, L. J.; Payne, J. T.; Lewis, J. C. Late-stage diversification of biologically active molecules via chemoenzymatic C-H functionalization ACS Catal. 2016, 6, 1451– 1454 DOI: 10.1021/acscatal.5b02558

12. Kolev, J. N.; O'Dwyer, K. M.; Jordan, C. T.; Fasan, R. Discovery of potent parthenolide-based antileukemic agents enabled by late-stage P450-mediated C-H functionalization ACS Chem. Biol. 2014, 9, 164– 173 DOI: 10.1021/cb400626w

13. Wang, H.; Negretti, S.; Knauff, A. R.; Montgomery, J. Exo-selective reductive macrocyclization of ynals Org. Lett. 2015, 17, 1493– 1496 DOI: 10.1021/acs.orglett.5b00381

14. Jackson, E. P.; Malik, H. A.; Sormunen, G. J.; Baxter, R. D.; Liu, P.; Wang, H.; Shareef, A.-R.; Montgomery, J. Mechanistic basis for regioselection and regiodivergence in nickel-catalyzed reductive couplings Acc. Chem. Res. 2015, 48, 1736– 1745 DOI: 10.1021/acs.accounts.5b00096

15. Michaudel, Q.; Journot, G.; Regueiro-Ren, A.; Goswami, A.; Guo, Z. W.; Tully, T. P.; Zou, L. F.; Ramabhadran, R. O.; Houk, K. N.; Baran, P. S. Improving physical properties via C-H oxidation: Chemical and enzymatic approaches Angew. Chem., Int. Ed. 2014, 53, 12091– 12096 DOI: 10.1002/anie.201407016

16. Wilde, N. C.; Isomura, M.; Mendoza, A.; Baran, P. S. TWo-phase synthesis of (−)-taxuyunnanine D J. Am. Chem. Soc. 2014, 136, 4909– 4912 DOI: 10.1021/ja501782r

17. Gormisky, P. E.; White, M. C. Catalyst-controlled aliphatic C-H oxidations with a predictive model for site-selectivity J. Am. Chem. Soc. 2013, 135, 14052– 14055 DOI: 10.1021/ja407388y

18. Chen, M. S.; White, M. C. Combined effects on selectivity in Fe-catalyzed methylene oxidation Science 2010, 327, 566– 571 DOI: 10.1126/science.1183602

19. Negretti, S.; Narayan, A. R. H.; Chiou, K. C.; Kells, P. M.; Stachowski, J. L.; Hansen, D. A.; Podust, L. M.; Montgomery, J.; Sherman, D. H. Directing group-controlled regioselectivity in an enzymatic C-H bond oxygenation J. Am. Chem. Soc. 2014, 136, 4901– 4904 DOI: 10.1021/ja5016052

20. Narayan, A. R. H.; Jimenez-Oses, G.; Liu, P.; Negretti, S.; Zhao, W. X.; Gilbert, M. M.; Ramabhadran, R. O.; Yang, Y. F.; Furan, L. R.; Li, Z.; Podust, L. M.; Montgomery, J.; Houk, K. N.; Sherman, D. H. Enzymatic hydroxylation of an unactivated methylene C-H bond guided by molecular dynamics simulations Nat. Chem. 2015, 7, 653– 660 DOI: 10.1038/nchem.2285

21. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes Science 2013, 339, 307– 310 DOI: 10.1126/science.1231434

22. Hyster, T. K.; Knorr, L.; Ward, T. R.; Rovis, T. Biotinylated Rh(III) complexes in engineered streptavidin for accelerated asymmetric C-H activation Science 2012, 338, 500– 503 DOI: 10.1126/science.1226132

23. Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y. E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. An artificial metalloenzyme with the kinetics of native enzymes Science 2016, 354, 102– 106 DOI: 10.1126/science.aah4427

24. Emmanuel, M. A.; Greenberg, N. R.; Oblinsky, D. G.; Hyster, T. K. Accessing non-natural reactivity by irradiating nicotinamide-dependent enzymes with light Nature 2016, 540, 414– 417 DOI: 10.1038/nature20569

25. Das, S.; Incarvito, C. D.; Crabtree, R. H.; Brudvig, G. W. Molecular recognition in the selective oxygenation of saturated C-H bonds by a dimanganese catalyst Science 2006, 312, 1941– 1943 DOI: 10.1126/science.1127899

26. Tang, R. Y.; Li, G.; Yu, J. Q. Conformation-induced remote meta-C-H activation of amines Nature 2014, 507, 215– 220 DOI: 10.1038/nature12963

27. Breslow, R. Biomimetic control of chemical selectivity Acc. Chem. Res. 1980, 13, 170– 177 DOI: 10.1021/ar50150a002

28. Breslow, R.; Baldwin, S.; Flechtner, T.; Kalicky, P.; Liu, S.; Washburn, W. Remote oxidation of steroids by photolysis of attached benzophenone groups J. Am. Chem. Soc. 1973, 95, 3251– 3262 DOI: 10.1021/ja00791a031

29. Li, S.; Chaulagain, M. R.; Knauff, A. R.; Podust, L. M.; Montgomery, J.; Sherman, D. H. Selective oxidation of carbolide C-H bonds by an engineered macrolide P450 mono-oxygenase Proc. Natl. Acad. Sci. U. S. A. 2009, 106, 18463– 18468 DOI: 10.1073/pnas.0907203106

30. Hein, J. E.; Fokin, V. V. Copper-catalyzed azide-alkyne cycloaddition (CuAAC) and beyond: New reactivity of copper(I) acetylides Chem. Soc. Rev. 2010, 39, 1302– 1315 DOI: 10.1039/b904091a

31. Boren, B. C.; Narayan, S.; Rasmussen, L. K.; Zhang, L.; Zhao, H. T.; Lin, Z. Y.; Jia, G. C.; Fokin, V. V. Ruthenium-catalyzed azide-alkyne cycloaddition: Scope and mechanism J. Am. Chem. Soc. 2008, 130, 8923– 8930 DOI: 10.1021/ja0749993

32. Li, S.; Ouellet, H.; Sherman, D. H.; Podust, L. M. Analysis of transient and catalytic desosamine-binding pockets in cytochrome P-450 pikC from streptomyces venezuelae J. Biol. Chem. 2009, 284, 5723– 5730 DOI: 10.1074/jbc.M807592200

33. Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2013.

34. (a) Becke, A. D. Phys. Rev. A 1988, 38, 3098– 3100. (b) Becke, A. D. J. Chem. Phys. 1993, 98, 5648– 5652. (c) Lee, C.; Yang, W.; Parr, R. G. Phys. Rev. B 1988, 37, 785– 789.

35. (a) Zhao, Y.; Truhlar, D. G. Phys. Chem. Chem. Phys. 2008, 10, 2813-2818. (b) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. J. Phys. Chem. B. 2011, 115, 14556-14562.

36. (a) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. J. Chem. Phys. 2010, 132, 154104: 1-19. (b) Grimme, S.; Ehrlich, S.; Goerigk, L. J. Comp. Chem. 2011, 32, 1456-1465.

37. (a) Barone, V.; Cossi, M. J. Phys. Chem. A 1998, 102, 1995-2001. (b) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. J. Comp. Chem. 2003, 24, 669-681.

38. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J. Chem. Theory Comput. 9, 3878–3888 (2013).

39. Case, D. A. et al. AMBER 12 (UCSF, 2012).

40. K. Shahrokh, A. Orendt, G. S. Yost, T. E. Cheatham, Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. J. Comput. Chem. 33, 119–133 (2012).

41. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of a general amber force field. J. Comput. Chem. 25, 1157–1174 (2004).

42. C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J. Phys. Chem. 97, 10269–10280 (1993).

43. (a)B. H. Besler, K. M. Merz, P. A. Kollman, Atomic charges derived from semiempirical methods. J. Comput. Chem. 11, 431–439 (1990). (b) U. C. Singh, P. A. Kollman, An approach to computing electrostatic charges for molecules. J. Comput. Chem. 5, 129–145 (1984).

44. Frisch, M. J. et al. Gaussian 09 (Gaussian, 2009).

45. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. 79, 926–935 (1983).

46. J. Wang, P. Cieplak, P. A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 21, 1049–1074 (2000).

47. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics. 98, 10089–10092 (1993).

Chapter 2. Influence of Water and Enzyme SpnF on the Dynamics and Energetics of the Ambimodal [6+4]/[4+2] Cycloaddition

2.1 Abstract

SpnF is the first monofunctional Diels–Alder/[6+4]-ase that catalyzes a reaction leading to both Diels–Alder and [6+4] adducts through a single transition state. The environment-perturbed transition-state sampling method has been developed to calculate free energies, kinetic isotope effects, and quasi-classical reaction trajectories of enzyme-catalyzed reactions and the uncatalyzed reaction in water. Energetics calculated in this way reproduce the experiment and show that the normal Diels–Alder transition state is stabilized by H bonds with water molecules, while the ambimodal transition state is favored in the enzyme SpnF by both intramolecular hydrogen bonding and hydrophobic binding. Molecular dynamics simulations show that trajectories passing through the ambimodal transition state bifurcate to the [6+4] adduct and the Diels–Alder adduct with a ratio of 1:1 in the gas phase, 1:1.6 in water, and 1:11 in the enzyme. This example shows how an enzyme acts on a vibrational time scale to steer post-transition state trajectories toward the Diels–Alder adduct.

2.2 Introduction

The enzyme SpnF catalyzes the transannular Diels–Alder (DA) reaction that converts 1 to 2 (Fig. 1) in the biosynthesis of spinosyn A, a major component of the insecticide spinosad [(1)]. Previous dynamics simulations of this reaction in the gas phase [(2)] show that the reaction occurs via an ambimodal [(3)] transition state (TS), and a path that bifurcates to form two products: the experimentally observed DA ([4+2]) adduct 2 and an unobserved [6+4] adduct 3. The initial product ratio of 2 to 3 in the gas phase was predicted to be 2:5. The transition state is highly asynchronous, and the first C–C bond is formed 140– 170 fs before the second [(2)]. The selectivity

of the reaction is governed not by TS partition functions but by dynamics that are sensitive to the environment [4, 5]. We have now explained how the intrinsic mechanism is modified, energetically and dynamically, by aqueous solution and by the SpnF enzyme. Inparticular, our developed method, environment-perturbed transition-state sampling, enables the discovery of single-molecule time-resolved reaction pathways in condensed media.

The wide applications of the DA reaction [6, 7] inspire the identification of natural enzymes that catalyze DA reactions (Diels–Alderases or DAases) [8]. While extensive surveys of secondarymetabolitesindicatethat hundredsofnaturalproducts are potentially biosynthesized by DAases [9], only a handful of purified enzymes have been demonstrated to catalyze DA reactions. These enzymes often catalyze other reactions, such as oxidations or degradations, thus leaving uncertain their specific influence on the DA reactions [10]. In 2011, Liu and coworkers discovered the first monofunctional DAase, SpnF, which catalyzes a DA reaction in the biosynthetic pathway of spinosyn A. Although the crystal structure of SpnF was reported in 2015 [11], the mechanism by which the enzyme catalyzes the DA reaction is still largely unknown.

Computational simulations have been employed to rationalize the reaction mechanism [12], but accurate mechanistic studies in solvent and in enzyme remain challenging [13], because these require both accurate description of interatomic interactions and extensive sampling of molecular conformations [14]. For instance, the recent study of AbyU-catalyzed DA reaction by Race and coworkers [15] has shown that the proposed DA reaction likely proceeds via a concerted, yet highly asynchronous transition state in the active site of the enzyme AbyU. However, the calculated potential of mean force (PMF) barrier using a hybrid quantum-mechanics classical molecular mechanics method severely underestimates the experimental barrier.

Additionally, as single-molecule experimental techniques to study reactions advance [16], mechanistic studies are no longer limited to the description of ensemble-average thermodynamics and kinetics. The explorations of chemical reactions now extend to the atomic level and femtosecond–picosecond time scale, which informs how atomic motions, conformational fluctuations, and solvent effects alter the reaction pathways of single molecules [17]. This accordingly calls for the development of computational tools that investigate the energetics and dynamics of single molecules in condensed media. We have developed an environment-perturbed transition-state sampling (EPTSS) method for individual trajectories. The EPTSS method has also been used for free-energy calculations and kinetic isotope effect calculations in solvents and enzymes. This enables the unraveling of the role of solvents and enzymes on control of single-molecule reaction pathways involving ambimodal transition states and potential energy surface (PES) bifurcations [5,18]. Quasi-classical trajectorieshavebeenemployed previously in the gas phase to understand and predict selectivities of reactions with post-TS bifurcations on the PES [5, 19– 21], such as for the cationic rearrangements occurring in terpene biosynthesis [22, 23]. We report how solvent and enzyme influence the dynamical behavior of a bifurcating reaction path in the SpnF-catalyzed DA reaction.

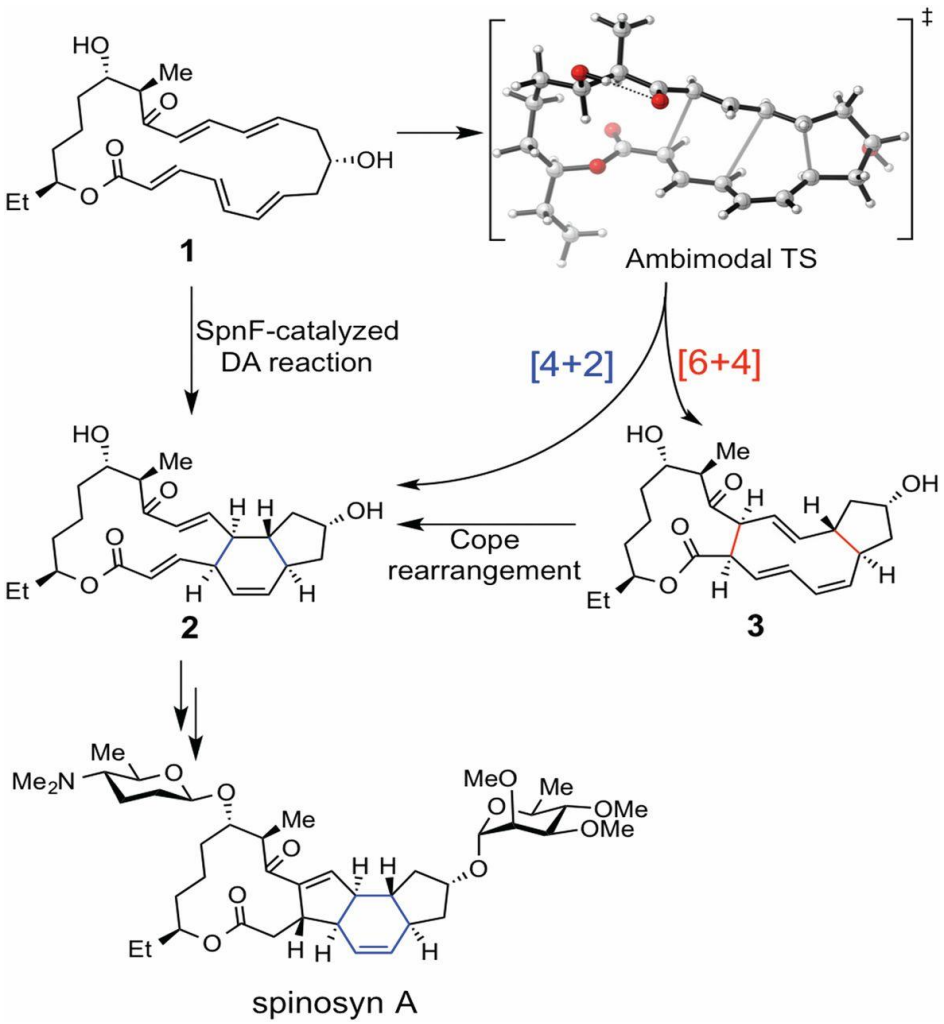


Figure 2.1. SpnF-catalyzed transannular DA reaction of 1 to form 2. This is a step in the biosynthesis of spinosyn A.

2.3 Computational Methods

The EPTSS method consists of four major steps. (i) Construction of the reaction/medium models. Initial structures of reactants and transition states were optimized in the gas phase at the M06-2X/6–31G(d) level of theory [39], using Gaussian 09 [40]. The optimized structures were then solvated in a water box using AmberTools 14 [41] for the aqueous system. For the reaction in enzyme, the substrate or TS were docked into the enzyme active site using AutoDock Vina. (ii) Configuration Sampling. Separate ensembles of R-A and R-B reactants were constructed by carrying out classical MD equilibrations in water and SpnF using MM for reactant and

environment. Classical MD was performed using Amber 14 on the substrate and the transition state for 10 ns in water and for 500 ns in SpnF. The FF99SBildn force field was used for protein residues. General Amber Force Field was used for reactant and transition structures. During the classical MD on the TS, restraining potentials of 500 kcal/mol/Å^2 were applied to the reaction coordinates in the TS. Snapshots (typically 100) of reactant and TS were sampled from production MD runs at 5-ps intervals in water and 1-ns intervals in SpnF. (iii) Free-energy calculations. To compute the QM/MM free-energy differences, we used a modified form of the ensemble-averaged variational transition state theory (EA-VTST) of Truhlar et al. [27, 42], which incorporates vibrational contributions to the PMF in condensed phase. For each snapshot, the coordinates of the environment (water or SpnF) were frozen, the geometry of 1 was optimized with QM/MM with M06-2X/6–31G(d) as QM, and the free energy was computed from harmonic vibrational partition functions at the stationary point. Free energies of TS-A and TS-B were computed similarly, using conventional TST instead of locating the TS variationally (VTST). This form of the theory is called EA-TST. In our previous study, the reaction-path VTST calculation is performed along the intrinsic reaction coordinate initiated from the ambimodal TS-A in the gas phase. The rate-determining VTS highly resembles the TS, which justifies the EA-TST we used here. In the TS equilibrations, reacting bond lengths were restrained as described in Methods. Free-energy barriers were computed as $\langle E(\text{TS-A}) - \langle E(\text{R}) \rangle$ and $\langle E(\text{TS-B}) \rangle - \langle E(\text{R}) \rangle$, where angle brackets indicate an ensemble average over 100 snapshots. Our free-energy scheme assumes that enzyme is well preorganized. The assumption has been justified in other studies [22]. On the other hand, for reactions without large charge separation, the preorganization energy is relatively trivial. The SpnF-catalyzed DA reaction was shown to involve very little charge separation by Medvedev et al. [27]. This also justifies our method used here. (iv) Reaction dynamics. Reactive trajectory

simulations were initiated from the random normal-mode sampled TS in water or in enzyme. Normal-mode sampling was conducted at 300 K for each sampled TS structure to obtain coordinates and momenta in a quasi-classical manner. These trajectories were propagated forward and backward for 150 fs each.

2.4 Results and Discussion

Using quantum mechanics/molecular mechanics (QM/MM) methods [24, 25], we have examined the effect of water and SpnF on the dynamics and initial product ratios of the transannular cycloadditions of 1. We extended our previously described solvent-perturbed TS sampling method [26] to carry out quasi-classical normal-mode sampling of the ensemble of TS in enzymes. The method is now called the EPTSS to reflect its generality.

Conformational Ensembles for Reactant, Ambimodal TS-A, and DA TS-B.

The substrate 1 is a conformationally flexible 22-membered polyene lactone. The conformation complexity was thoroughly analyzed by Medvedev et al. [27]. We find that the conformers of 1 consist mainly of two clusters represented in Fig. 2 by the gasphase conformers R-A and R-B. They differ mainly in the presence or absence of an intramolecular hydrogen bond. This is reflected in the highlighted dihedral angle Φ. R-A has an intramolecular hydrogen bond (H-bond) and R-B does not (Fig.3). TS-A is more stable than TS-B by 3.3 kcal/mol in the gas phase.

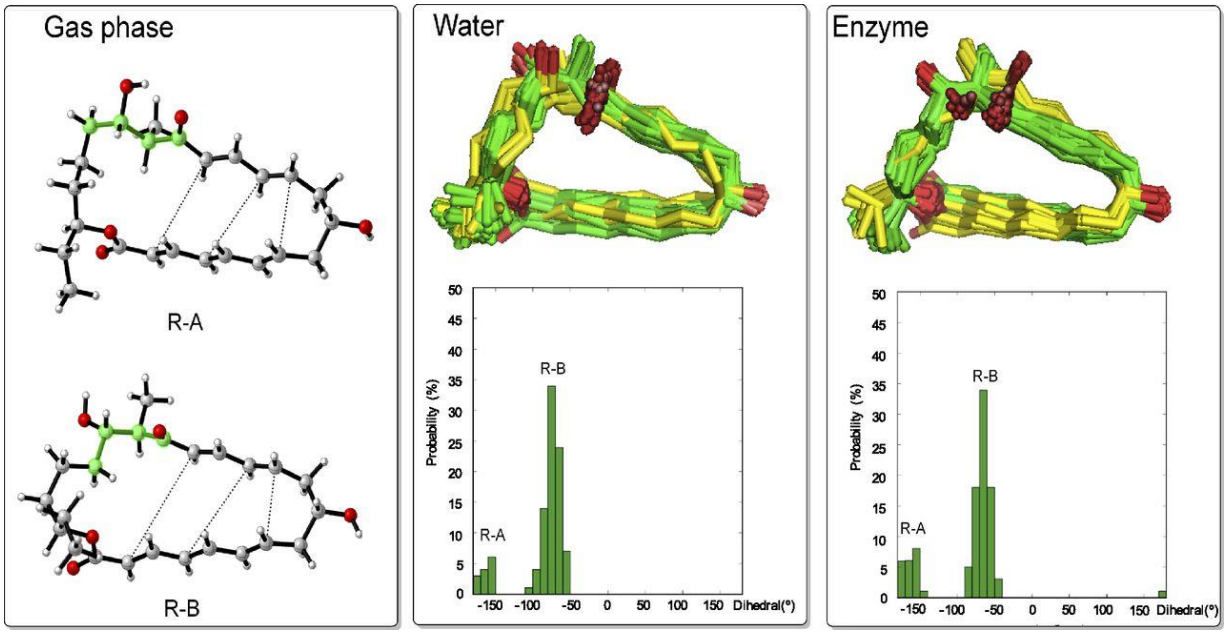


Figure 2.2. Ensembles of reactants from classical MD in water and in the enzyme. Water molecules and enzyme residues are not displayed. R-A and R-B are two representative conformations optimized in the gas phase. The dihedral angle used for discriminating between the two conformations is highlighted.

Classical molecular dynamics (MD) equilibrations in water and SpnF predict an 80:20 equilibrium mixture of R-B:R-A in both water and SpnF. This contrasts with the gas phase, where R-A is 5.2 kcal/mol below R-B, largely due to the intramolecular H bond. In water and in SpnF, hydrogen bonds to the carbonyl of R-B (Q148, I40 in SpnF) stabilize R-B.

MD simulations with restrained TS [28] geometries were conducted separately on both TS conformers in water and enzyme SpnF. Ensembles for ambimodal TS-A and for DA TS-B were constructed separately by taking 100 snapshots of the solvent box containing the constrained TSs with a 100-ps interval in water and a 1-ns interval in the enzyme. For each ensemble, subsequent QM/MM calculations were conducted to optimize the reactant and the ambimodal TS-A or DA TS-B in the perturbing environment. The free energies of each were computed by averaging over the entire ensemble.

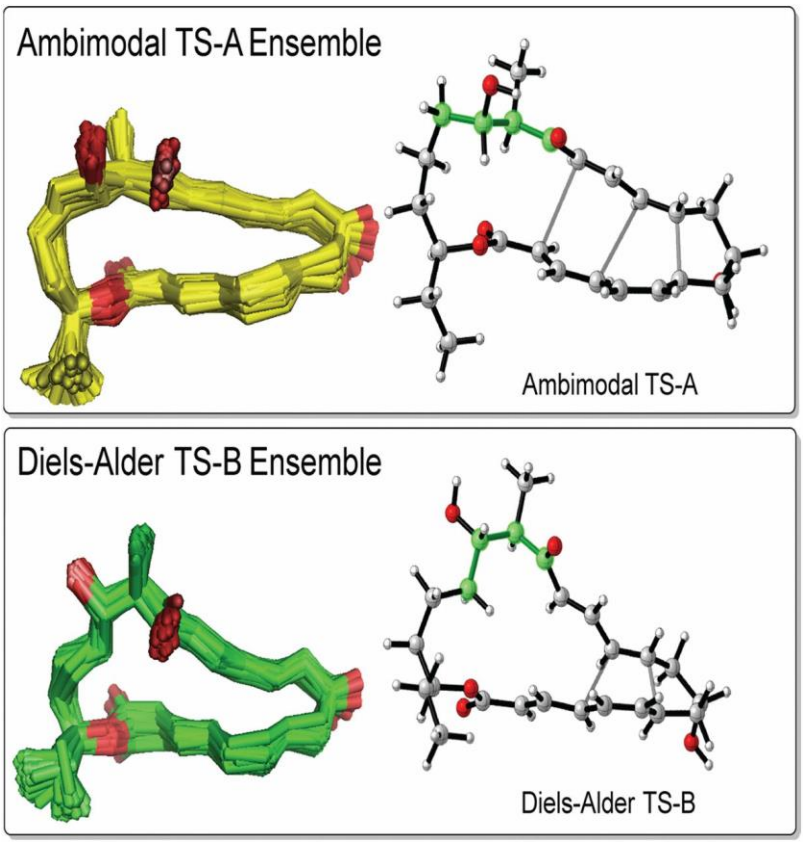


Figure 2.3. Ensembles of ambimodal TS-A and DA TS-B in enzyme. Ambimodal TS-A and DA TS-B are representative conformations for their corresponding ensembles. The dihedral angle used for discriminating two conformations is highlighted.

Computed Free Energies of Activation.

Table 1 shows the averaged free-energy barriers for ambimodal TS-A and DA TS-B in water and in the enzyme SpnF. Averages for different numbers of snapshots are computed. The averages based on a small number of snapshots have large SEs. The SEs decrease, and the energetics converge after including a large number of snapshots. Sufficient conformational sampling is essential for the free-energy calculations of reactions in condensed media [29, 30].

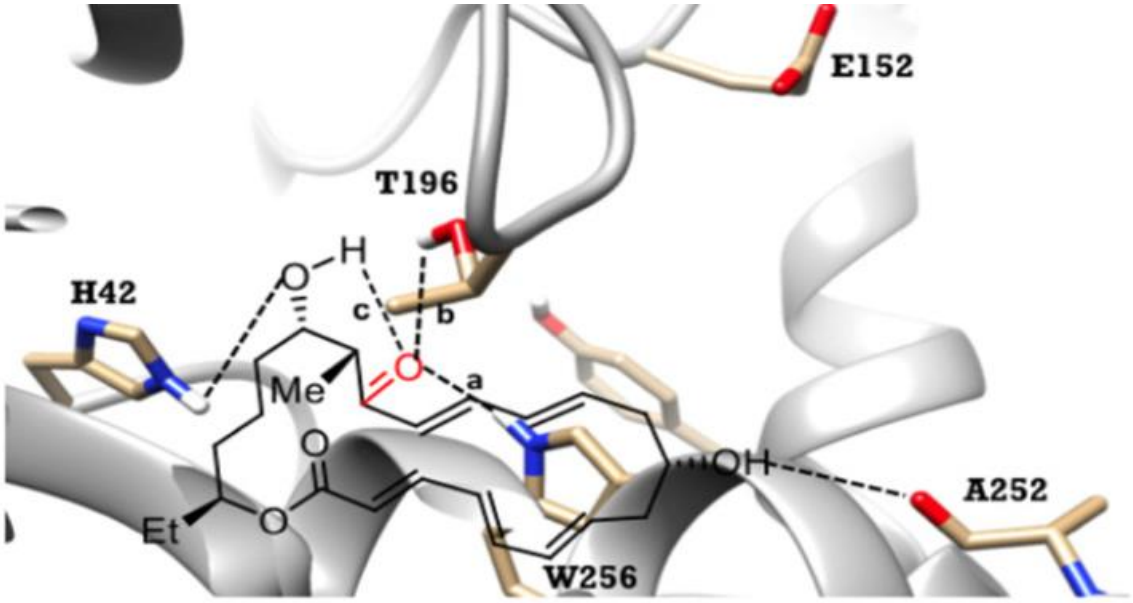
Table 2.1. Free-energy barriers for cycloadditions of 1 in water and SpnF enzyme

Medium	Ambimodal TS-A, kcal/mol	DA TS-B, kcal/mol	Experiment, kcal/mol*	Experiment, min^{-1}
Water	27.8 ± 1.9†	22.9 ± 2.1†	22.0	0.288 ± 0.0004
Enzyme	18.8 ± 1.0†	23.4 ± 1.1†	18.3	14 ± 1.6

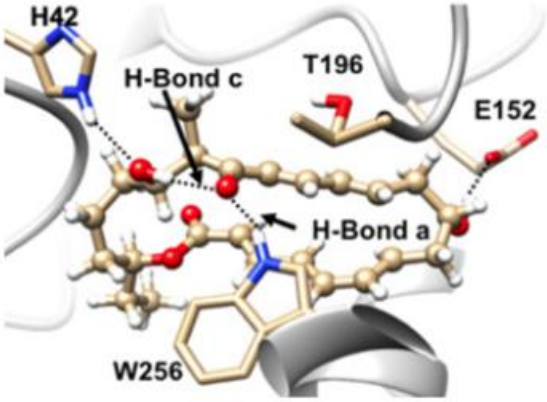
*Barrier obtained from experimental rate via the Eyring equation.
†SE after averaging over 100 snapshots.

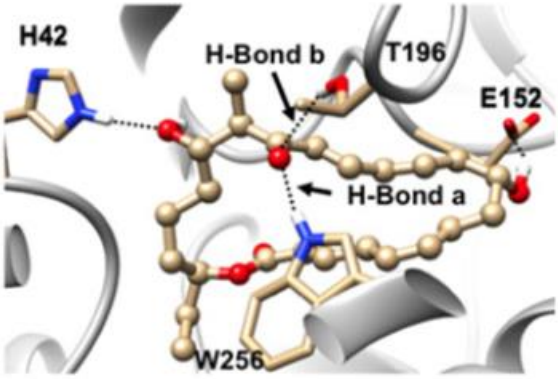
The DA TS-B is preferred in water, while the ambimodal TSAisfavored in the enzyme. This is a dramatic difference between the reaction in water and in SpnF enzyme. The computed barriers are consistent with the experimental values of 22.0 ± 0.01 kcal/mol in water and 18.3 ± 0.1 kcal/mol in SpnF, calculated from rates measured at 25 °C (1). The QM/MM umbrella sampling calculation has been recently performed by Zheng and Thiel [31]. These simulations show a free-energy barrier of 22 kal/mol for D TS-B in the SpnF enzyme. This is also consistent with our result. We have also computed the secondary kinetic isotopic effect (KIE) for the reaction in water and in enzyme, and compared with the experimental results [32]. Given experimental errors and statistical errors due to our limited computational sampling, the results are in reasonable agreement. The EPTSS KIEs are averaged over the entire TS ensemble. Both experiment and theory establish a significant inverse kH/kD at C7–C11, in the ratedetermining TSs in both water and enzyme, indicating bond formation. Nearly negligible KIE are found experimentally and computationally at C4, C12, C2, and C14. The only exception is for C4, where both experiment and theory predict no KIE in water, but small inverse in the enzyme. These results support that TS-B is favored in water, while TS-A in preferred in the enzyme.

Table 2.2. Percentage of structures that have H-bond interactions with enzyme residues in ensembles of reactant, TS-A, and TS-B after QM/MM optimizations. A H bond is defined as having a H–O bond length shorter than 2.25 Å and an O–H–O bond angle greater than 150°



Ambimodal TS-A Diels-Alder TS-B





	H-bond *a*, %	H-bond *b*, %	H-bond *c*, %
Ambimodal TS-A	86	0	99
DA TS-B	87	21	0
Reactant	61	0	9

The elucidation of hydrogen-bonding patterns gives insight into the origins of catalysis. In water, TS-B is better stabilized than TS-A by forming more intermolecular H bonds with water molecules. TS-A already has an intramolecular H bond and forms fewer H bonds with water. Additionally, TS-B is more polarized than TS-A.

The hydrogen bonds measured in the reactant and two conformers of the TS in the enzyme are shown in Table 2. H bonds between substrate and H42, E152 (to the TS), and A252 (to the reactant) stabilize the binding configuration of the TS and the reactant. Two typical snapshots are also shown for ambimodal TS-A and DA TS-B with H bonding to their surrounding residues. H bonds between the residues and the highlighted C=O decrease the energy barriers of the cycloadditions [33].

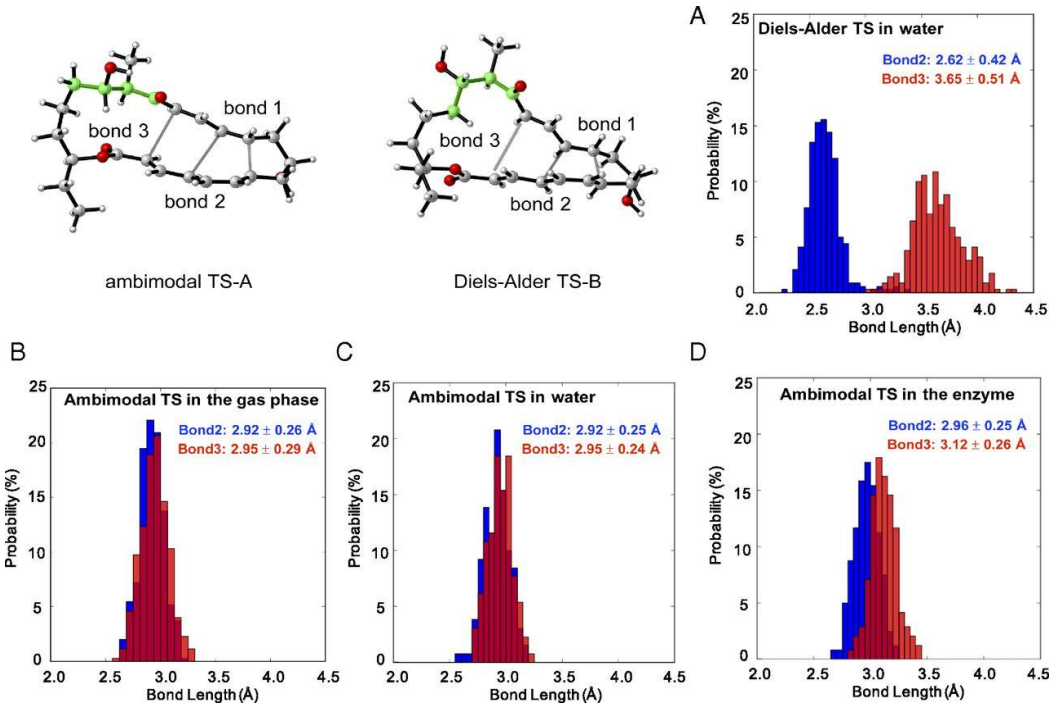


Figure 2.4. Distribution of bonds 2 and 3 in 240 TS geometries for (A) DA TS-B in water, (B) ambimodal-A TS in the gas phase, (C) ambimodal TS-A in water, and (D) ambimodal TS-A in the enzyme. Bond 2 in blue leads to the [4+2] adduct, while bond 3 in red leads to the [6+4] adduct. For A, C, and D, TS geometries were sampled by using normal-mode sampling on 60 transition structures optimized in various snapshots of enzyme or of water.

In 100 snapshots in the enzyme, H bonds to W256 are present in 61% of the reactant snapshots, and in 86% and 87% of the ambimodal TS and DA TS, respectively. From reactant to TS, the ambimodal TS-A experiences an enhancement of H-bond c (intramolecular H bond) from 9 to 99%, while DA TS-B shows only a small increase in H-bond b (toT195) percentage from 0 to

21%. This explains in part why the ambimodal TS is more favorable in the enzyme. The protein pocket is highly hydrophobic, and only 0.2 water molecules are detected that bind to the hydroxyl and carbonyl group of the DA TS-B during the MD. The DA TS-B in the enzyme is not well stabilized by intermolecular H bonds provided by water or protein residues. This is in direct contrast to the previous computation by Gordeev and Ananikov [34] based on a theozyme model where all polar functional groups of the substrate bind to the amino acids.

Reaction Dynamics Simulations.

Enzymes can bind reactants and TS by both electrostatic and hydrophobic interactions. These interactions may favor certain TS geometries, and influence reaction dynamics as well. The role of dynamics on catalysis by enzymes has been a significant topic of discussion ever since Falzone, Wright, and Benkovic reported correlations between dynamical motions of remote residues and changes in rate by mutation of these residues [35]. After considerable debate, the current consensus is that relatively slow (millisecond–microsecond) motions of loops and remote residues do indeed alter the structure of the active site and the rate of reaction [36], but fast (picosecond–femtosecond) couplings of vibrational motions to motions along the reaction coordinate have little or no influence on the rate of reactions. We note that residues in contact with the reactant as they vibrate through the transition zone can influence energetics, and the femto second motions of these residues can be coupled to the reaction event, while even nanosecond loop motion is too slow to influence reaction dynamics.

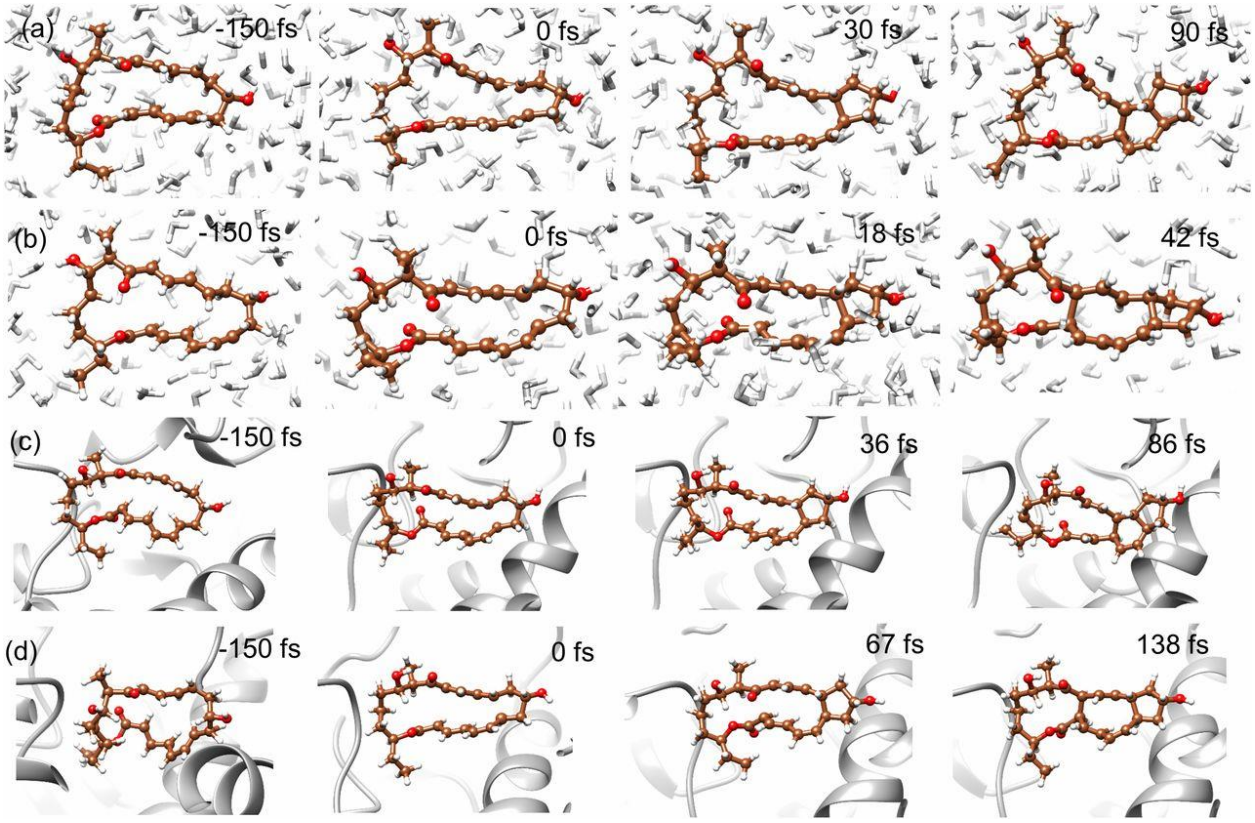


Figure 2.5. Typical trajectories for the formation of (A) [4+2] adduct in water, (B) [6+4] adduct in water, (C) [4+2] adduct in enzyme, (D) [6+4] adduct in enzyme. We define 1.6 Å as the criterion for C–C bond formation.

Fig. 4 shows the distribution of sampled geometries for the DA TS-B in water (Fig. 4A), and the ambimodal TS-A in the gas phase (Fig. 4B), in water (Fig. 4C) and in SpnF enzyme (Fig. 4D). Bonds 2 and 3 are formed in the [4+2] and the [6+4] adducts, respectively. For DA TS-B in water, the distribution of bond 2 is 2.62 ± 0.42 Å, and that of bond 3 is 3.65 ± 0.51 Å. The two bonds on average differ by ∼1 Å. The formation of the [4+2] adduct is almost always observed in dynamics from this TS (vide infra). In contrast, the ambimodal TS-A in enzyme shows that bonds 2 and 3 are much closer in length with 2.96 ± 0.25 Å for bond 2 and 3.12 ± 0.26 Å for bond 3. In the gas phase, the difference between bonds 2 and 3 is further narrowed, making the formation of [6+4] adduct more likely. These results indicate that the nature of the TS varies in the gas, water,

and enzyme environment. This is a result of the average dynamical environment and alters the TS geometries.

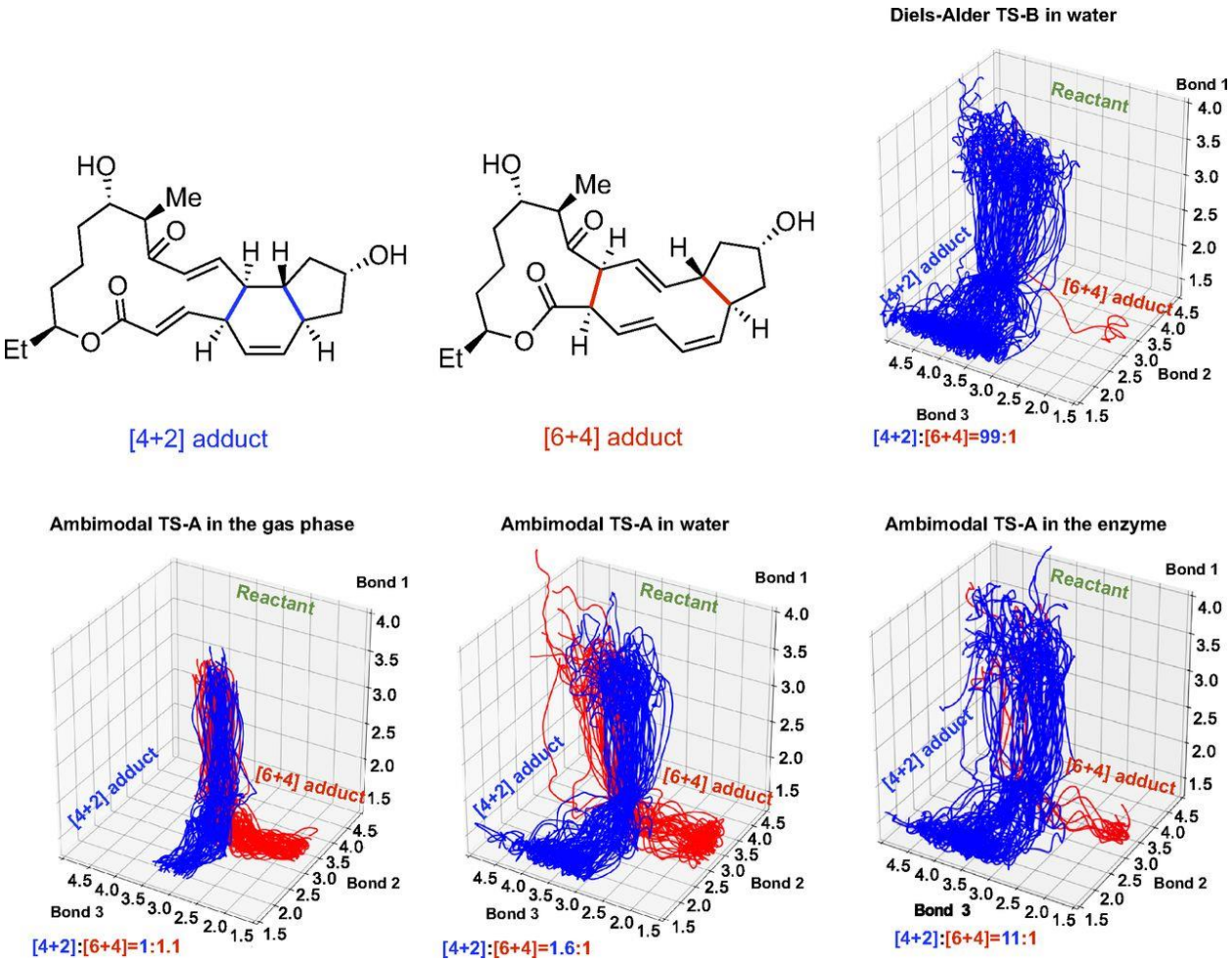


Figure 2.6. Distributions of reactive trajectories initiated from DA TS in water, ambimodal TS in the gas phase, in water, and in enzyme. One hundred randomly chosen trajectories were plotted in each case. Trajectories leading to [4+2] adduct are shown in blue, and those leading to [6+4] adduct are shown in red.

Reaction dynamics simulations are essential to elucidate how these environment perturbations change the pathways that trajectories have to take to achieve product formation. Reaction dynamics trajectories were initiated from TS geometries after normal-mode sampling to explore the dynamics of product formation. Each reactive trajectory was propagated for 300 fs using the QM/MM method.

Fig. 5 displays typical trajectories propagated in water and in the enzyme. The third and fourth panels in each row indicate the time atwhich bond 1and either bonds 2 or 3 are formed, defined as achieving a distance of 1.6 Å [37]. (Bonds 1, 2, and 3 are labeled in Fig. 4.) Bond 1 forms in both adducts. Bond 2 forms in the [4+2] adduct, while bond 3 forms in the [6+4] adduct. Fig. 5 A and B shows the production of [4+2] adduct and [6+4] adduct, respectively. In Fig. 5A, bond 1 forms at 30 fs, then bond 2 forms at 90 fs. In Fig. 5B, bond 1 forms at 18 fs, then bond 3 forms at 42 fs. Likewise, Fig. 5 C and D represents the formation of [4+2] adduct and [6+4] adduct, respectively, in the enzyme, with two bonds forming at 36 and 86 fs for the formation of the [4+2] adduct, and at 67 and 138 fs for the formation of the [6+4] adduct.

Fig. 6 shows overlays of 100 trajectories plotted according to lengths of forming bonds 1, 2, and 3. The ratios of [4+2]:[6+4] products are also presented. In water, the DA TS leads almost exclusively to DA adduct, with only 1 out of 100 trajectories leading to a [6+4] adduct. For trajectories passing through the ambimodal TS, bifurcation to form both products is observed; [4+2]:[6+4] ratios of 11:1 in enzyme, 1.6:1 in water, and 1:1.1 in the gas phase are completed. The selectivity for [4+2] and [6+4] adduct results from the competition between the formation of bond2 and bond 3. From gas phase to water, a trend toward [4+2] adduct was observed, even though the distributions of the transition state geometries forambimodal TS-A are very similar inboth media. The dipole moment of the [4+2] adduct is 4.74 debye, while that of the [6+4] adduct is 3.46 debye, and the [4+2] adduct is better stabilized by polar solvent. In addition, trajectories were propagated in implicit solvents [solvation model density (SMD) model [38]] to explore how the polarity of the medium influences the product distribution. The [4+2]:[6+4] ratio is 1:1 in implicit hexane and 1.5:1 in implicit water. This further supports that increase in solvent polarity promotes the formation of the [4+2] adduct over the [6+4] adduct.

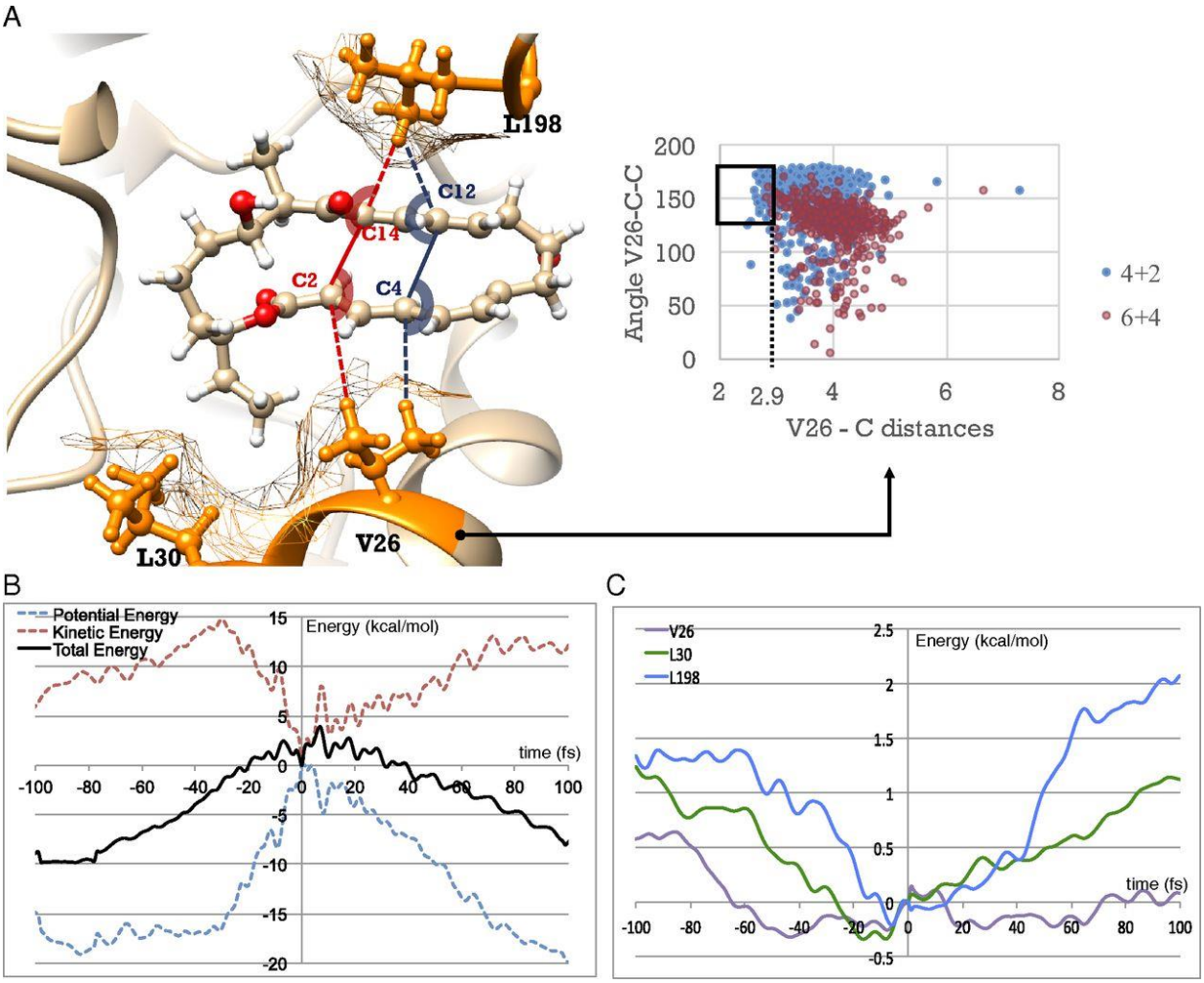


Figure 2.6. (A) Hydrophobic interactions in SpnF–TS complex. The distance is measured from the closest H on V26 or L30 (depending which residue is closer in the snapshot) to C4 or C2, and the angle is the closest H–C4–C14 or H–C2–C14. Framed are the strong hydrophobic contacts, defined as the angle larger than 150° and the distance shorter than 2.9 Å. (B) Average of potential energy, kinetic energy, and total energy of the substrate versus time. (C) Average of kinetic energy for residues V26, L30, and L198 versus time. Energies are averaged from 159 trajectories leading to the [4+2] adduct. The energy at 0 fs are set as zero.

Fig. 7A shows how the hydrophobic residue valine V26 influences the formation of the two adducts through interactions with bond 2 (C4–C12) and bond 3 (C2–C14). The distribution of hydrophobic contacts shows a stronger compression on bond 2 than bond 3, as framed on the graph. This leads to relatively larger difference of lengths of partial bond 2 and 3 in the ambimodal TS in the enzyme compared with the gas phase and in water, which favors the formation of the

[4+2] adduct. Leucine L198, the other residue nearby the TS, involves similar hydrophobic interactions with the two bonds.

We analyzed the evolution of averaged potential energy (PE), kinetic energy (KE), and total energy (KE+PE) for 159 reactive trajectories leading to the [4+2] adduct (Fig. 7B). From −100 to −30 fs, KE increases by about 10 kcal/mol while PE remains relatively unchanged. This is a phase of thermal activation of the substrate to achieve reactive conformations. Subsequently, 15 kcal/mol KE is transferred into PE to form the TS at 0 fs. After passing the TS, the product is gradually formed with KE increasing and PE decreasing. The formation of bond 2 takes relatively longer time than bond 1, and PE and KE change smoothly. Fig. 7C displays the decreases of the kinetic energies for the three hydrophobic residues (V26, L30, and L198) surrounding the reacting molecule from −100 to 0 fs. Their kinetic energy in total decreases by ∼3 kcal/mol. We found that the KEs for polar residues that form H bonds to the substrate are always within 0.5 kcal/mol. These results indicate a kinetic energy transfer from the hydrophobic residues to the reacting substrate by vibrational collisions on a femtosecond time scale.

2.5 Conclusion

The EPTSS method [(26)] has been developed and applied to compute free-energy barriers and quasi-classical dynamics in aqueous solution and in the enzyme SpnF. Using this method, we investigated medium effects and femtosecond dynamics of SpnFcatalyzed transannular DA reactions. Two distinct TS were found for the reaction in water and in the enzyme. The computed reaction barriers are in good agreement with the experimentally measured rates of reaction. Important residues that contribute to the overall catalytic process and product distributions were identified. Our study shows how water and enzyme residues influence the rates of reaction and the distributions of products in a dynamically controlled ambimodal reaction mechanism.

2.6 References

1. Kim HJ, Ruszczycky MW, Choi S-H, Liu Y-N, Liu HW (2011) Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. Nature 473:109–112.

2. Patel A, et al. (2016) Dynamically complex [6+4] and [4+2] cycloadditions in the biosynthesis of spinosyn A. J Am Chem Soc 138:3631–3634.

3. Yu P, Patel A, Houk KN (2015) Transannular [6 + 4] and ambimodal cycloaddition in the biosynthesis of heronamide A. J Am Chem Soc 137:13518–13523.

4. Carpenter BK (1995) Dynamic matching: The cause of inversion of configuration in the [1, 3] sigmatropic migration? J Am Chem Soc 117:6336–6344.

5. Lourderaj U, Park K, Hase WL (2008) Classical trajectory simulations of post-transition state dynamics. Int Rev Phys Chem 27:361–403.

6. Diels O, Alder K (1928) Synthesen in der hydroaromatischen reihe. Justus Liebigs Ann Chem 460:98–122.

7. Nicolaou KC, Snyder SA, Montagnon T, Vassilikogiannakis G (2002) The Diels–Alder reaction in total synthesis. Angew Chem Int Ed Engl 41:1668–1698.

8. Jeon BS, Wang S-A, Ruszczycky MW, Liu HW (2017) Natural [4 + 2]-cyclases. Chem Rev 117:5367–5388.

9. Oikawa H, Tokiwano T (2004) Enzymatic catalysis of the Diels-Alder reaction in the biosynthesis of natural products. Nat Prod Rep 21:321–352.

10. Minami A, Oikawa H (2016) Recent advances of Diels-Alderases involved in natural product biosynthesis. J Antibiot (Tokyo) 69:500–506.

11. Fage CD, et al. (2015) The structure of SpnF, a standalone enzyme that catalyzes [4 + 2] cycloaddition. Nat Chem Biol 11:256–258.

12. Cheng GJ, Zhang X, Chung LW, Xu L, Wu YD (2015) Computational organic chemistry: Bridging theory and experiment in establishing the mechanisms of chemical reactions. J Am Chem Soc 137:1706–1725.

13. Houk KN, Liu F (2017) Holy grails for computational organic chemistry and biochemistry. Acc Chem Res 50:539–543.

14. Lu X, et al. (2016) QM/MM free energy simulations: Recent progress and challenges. Mol Simul 42:1056–1078.

15. Byrne MJ, et al. (2016) The catalytic mechanism of a natural Diels–Alderase revealed in molecular detail. J Am Chem Soc 138:6095–6098.

16. Kneipp K, et al. (1997) Single molecule detection using surface-enhanced Raman scattering (SERS). Phys Rev Lett 78:1667–1670.

17. Lu HP, Xun L, Xie XS (1998) Single-molecule enzymatic dynamics. Science 282:1877–1882.

18. Ess DH, et al. (2008) Bifurcations on potential energy surfaces of organic reactions. Angew Chem Int Ed Engl 47:7592–7601.

19. Yang Z, Yu P, Houk KN (2016) Molecular dynamics of dimethyldioxirane C–H oxidation. J Am Chem Soc 138:4237–4242.

20. Nieves-Quinones Y, Singleton DA (2016) Dynamics and the regiochemistry of nitration of toluene. J Am Chem Soc 138:15167–15176.

21. Yu P, et al. (2017) Mechanisms and origins of periselectivity of the ambimodal [6 + 4] cycloadditions of tropone to dimethylfulvene. J Am Chem Soc 139:8251–8258.

22. Hong YJ, Tantillo DJ (2009) A potential energy surface bifurcation in terpene biosynthesis. Nat Chem 1:384–389.

23. Major DT, Weitman M (2012) Electrostatically guided dynamics–The root of fidelity in a promiscuous terpene synthase? J Am Chem Soc 134:19454–19462.

24. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol 103:227–249.

25. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. Angew Chem Int Ed Engl 48:1198–1229.

26. Yang Z, Doubleday C, Houk KN (2015) QM/MM protocol for direct molecular dynamics of chemical reactions in solution: The water-accelerated Diels–Alder reaction. J Chem Theory Comput 11:5606–5612.

27. Medvedev MG, et al. (2017) Quantifying possible routes for SpnF-catalyzed formal Diels−Alder cycloaddition. J Am Chem Soc 139:3942–3945.

28. Noey EL, et al. (2015) Origins of stereoselectivity in evolved ketoreductases. Proc Natl Acad Sci USA 112:E7065–E7072.

29. Lin H, Truhlar DG (2007) QM/MM: What have we learned, where are we, and where do we go from here? Theor Chem Acc 117:185–199.

30. Masgrau L, Truhlar DG (2015) The importance of ensemble averaging in enzyme kinetics. Acc Chem Res 48:431–438.

31. Zheng Y, Thiel W (2017) Computational insights into an enzyme-catalyzed [4+2] cycloaddition. J Org Chem 82:13563–13571.

32. Jeon BS, et al. (2017) Investigation of the mechanism of the SpnF-catalyzed [4+2]-cycloaddition reaction in the biosynthesis of spinosyn A. Proc Natl Acad Sci USA 114:10408–10413.

33. Blake JF, Jorgensen WL (1991) Solvent effects on a Diels–Alder reaction from computer simulations. J Am Chem Soc 113:7430–7432.

34. Gordeev EG, Ananikov VP (2015) Computational study of a model system of enzyme-mediated [4+2] cycloaddition reaction. PLoS One 10:e0119984.

35. Falzone CJ, Wright PE, Benkovic SJ (1994) Dynamics of a flexible loop in dihydrofolate reductase from Escherichia coli and its implication for catalysis. Biochemistry 33:439–442.

36. Agarwal PK, Billeter SR, Rajagopalan PT, Benkovic SJ, Hammes-Schiffer S (2002) Network of coupled promoting motions in enzyme catalysis. Proc Natl Acad Sci USA 99:2794–2799.

37. Black K, Liu P, Xu L, Doubleday C, Houk KN (2012) Dynamics, transition states, and timing of bond formation in Diels-Alder reactions. Proc Natl Acad Sci USA 109:12860–12865.

38. Marenich AV, Cramer CJ, Truhlar DG (2009) Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J Phys Chem B 113:6378–6396.

39. Zhao Y, Truhlar DG (2008) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theor Chem Acc 120:215–241.

40. Frisch MJ, et al. (2009) Gaussian 09 (Gaussian, Inc. Wallingford, CT).

41. Case DA, et al. (2014) Amber14 (University of California, San Francisco).

42. Truhlar DG, et al. (2002) The incorporation of quantum effects in enzyme kinetics modeling. Acc Chem Res 35:341–349.

Chapter 3. Structural Basis of the Cope Rearrangement and Cyclization in Hapalindole Biogenesis

3.1 Abstract

Hapalindole alkaloids are a structurally diverse class of cyanobacterial natural products defined by their varied polycyclic ring systems and diverse biological activities. These complex metabolites are generated from a common biosynthetic intermediate by the Stig cyclases in three mechanistic steps: a rare Cope rearrangement, 6-*exo*-*trig* cyclization, and electrophilic aromatic substitution. Here we report the structure of HpiC1, a Stig cyclase that catalyzes the formation of 12-*epi*-hapalindole U in vitro. The 1.5-Å structure revealed a dimeric assembly with two calcium ions per monomer and with the active sites located at the distal ends of the protein dimer. Mutational analysis and computational methods uncovered key residues for an acid-catalyzed [3,3]-sigmatropic rearrangement, as well as specific determinants that control the position of terminal electrophilic aromatic substitution, leading to a switch from hapalindole to fischerindole alkaloids.

3.2 Introduction

The hapalindole family of alkaloids are a large and structurally diverse class of natural products from cyanobacteria of the order *Stigonematales*[1]. These metabolites are active against a broad range of targets, including antibacterial, antifungal, insecticidal, and antimitotic activities[2,3,4,5,6,7]. Each member is classified as a hapalindole, ambiguine, fischerindole, or welwitindolinone on the basis of its core ring system, and these have been the subject of various total syntheses due to their structural complexity and unique biological properties[1]. Until recently, comparatively little was known regarding the biogenesis of these alkaloids and, in particular, the construction of the tetracyclic core ring system.

Initial reports demonstrated that hapalindoles are derived from *cis*-indole isonitrile and geranyl pyrophosphate (GPP)[8,9,10], but the biogenesis of the polycyclic ring systems remained elusive. We recently identified an unexpected biosynthetic intermediate, 1, which undergoes a Cope rearrangement that is followed by a cyclization cascade to generate 12-*epi*-hapalindole U (5) (Fig. 1)[11]. The Cope rearrangement is a [3,3]-sigmatropic rearrangement that proceeds through a cyclic transition state[12]. Although this pericyclic reaction is prevalent in organic synthesis[13], it has rarely been identified as a biosynthetic transformation[14,15,16]. The biosynthesis of 5 was proposed to proceed through a three-part reaction mechanism: (i) Cope rearrangement of 1 to generate intermediate 3, which sets the stereochemistry at positions C11 and C12; (ii) 6-*exo*-*trig* cyclization of 3 to intermediate 4, which sets the stereochemistry at positions C10 and C15; and (iii) electrophilic aromatic substitution of 4 to yield 5 upon deprotonation (Fig. 1). This discovery was expanded to include several Stig cyclases and revealed that the variant configurations observed in this class of alkaloids are generated from the central biosynthetic intermediate 1, which is transformed to products in a regio- and stereospecific fashion by various members of the cyclase family (Fig. 1)[17,18,19]. Notably, FimC5, which yields the major product 12-*epi*-fischerindole U (6), generates a product that has a stereochemical configuration identical to that of 5 at C10, C11, C12, and C15, but differs in the site of C-ring attachment to the indole moiety. This indicates that the enzymes direct the site of terminal electrophilic aromatic substitution, thereby providing the regiochemical control that differentiates the hapalindoles and fischerindoles. Both HpiC1 and FimC5 produce low levels of tricyclic 12-*epi*-hapalindole C (7), a shunt metabolite that is not converted to its tetracyclic congeners by either enzyme. Thus, biogenesis of hapalindole-type metabolites includes a fascinating mechanistic puzzle in regard to how homologous Stig cyclases

maintain stereochemical and regiochemical control at each of these biosynthetic steps in the formation of varied hapalindoles and fischerindoles.

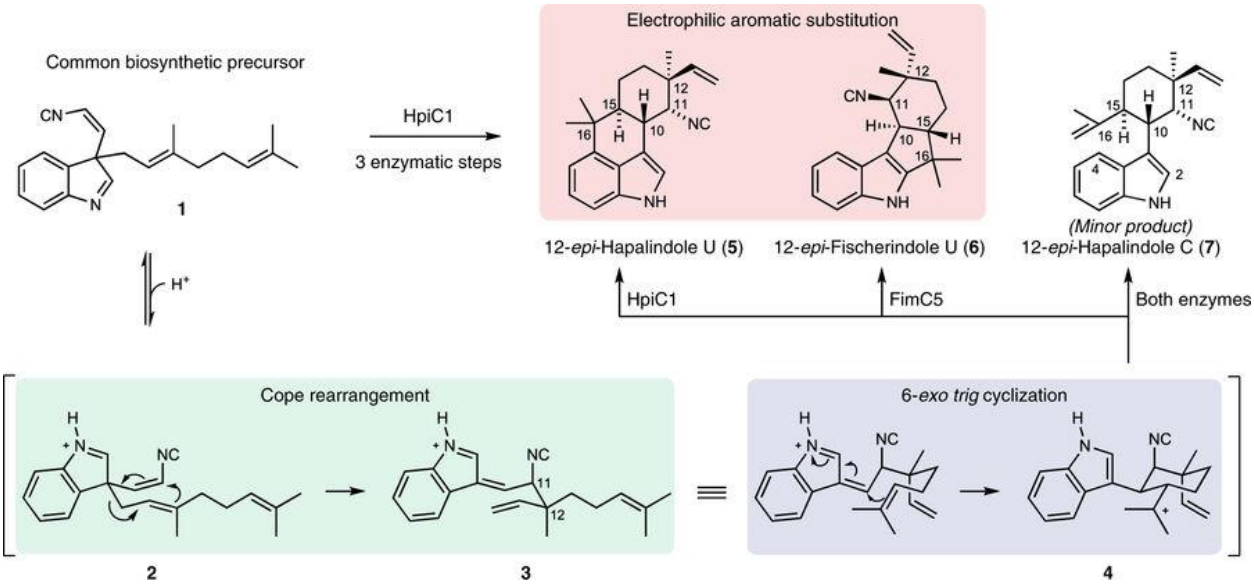


Figure 3.1. The hapalindole and fischerindole core ring systems arise from the common biosynthetic intermediate 1. Stig cyclases catalyze a Cope rearrangement and subsequent cyclization to generate tetracyclic products and trace levels of tricyclic shunt products. HpiC1 catalyzes formation of 5, while FimC5 catalyzes formation of 6, with identical stereochemistry at C11 and C12 but different C-ring regiochemistry.

In this study, we describe the molecular basis for the Stig cyclase's ability to control three reactions. As these were originally annotated as unknown proteins[8], no information was available regarding the structure of this new type of biosynthetic enzyme. We describe herein the first crystal structure of a Stig cyclase, HpiC1, and show through a mutational analysis, which localizes its active site, the ability to reconfigure its metabolite profile. Density functional theory (DFT) calculations on the mechanism and molecular dynamics simulations provide detailed information on the enzymatic processes that control product formation. These data provide compelling insights into the mechanism of Cope rearrangement, 6-*exo*-*trig* cyclization and electrophilic aromatic substitution for this class of natural products.

3.3 Methods

Cloning and mutagenesis of HpiC1 and FimC5

HpiC1 and FimC5 were cloned into pET28 (Novagen) from codon-optimized synthetic genes (IDT gBlocks) with their N-terminal leader peptides truncated17. Site-directed mutagenesis of HpiC1 and FimC5 was performed using a single-primer method based on "Quikchange" mutagenesis (Agilent Genomics). Mutagenic primer sequences are listed in Supplementary Table 2. All mutations were verified by DNA sequencing at the University of Michigan DNA Sequencing Core.

Expression of HpiC1 and FimC5 proteins

HpiC1 and FimC5 and their corresponding active site mutants were overexpressed in Escherichia coli strain BL21(DE3). Cultures from a single colony were used to inoculate 1.5 L terrific broth (TB) supplemented with 50 μg/mL kanamycin. Expression was induced with 0.7 mM isopropyl-β-d-thiogalactopyranoside when cultures reached OD600 ~1.0. After 20 h induction at 18 °C the cells were harvested by centrifugation and stored at −80 °C.

Expression of HpiC1 W73M/K132M selenomethionine derivative

An initial challenge involved the lack of native methionine residues in HpiC1. Therefore, a series of mutants containing methionine substitutions were screened for crystallization. Sites for substitution were selected based on positions containing a native methionine in sequence comparisons with other Stig cyclases. Selenomethionine (SeMet) HpiC1 W73M/K132M was produced by metabolic inhibition31. Briefly, freshly transformed BL21(DE3) cells harboring the hpiC1 gene on pET28 were used to inoculate 3 L M9 minimal medium supplemented with 50 μg/mL kanamycin. An amino acid cocktail containing l-selenomethionine was added when the cells reached OD600 = 1.0. The cells were cooled to 18 °C and shaken for 30 min before induction

with 0.7 mM isopropyl-β-d-thiogalactopyranoside. After 20 h induction, the cells were harvested by centrifugation and stored at −80 °C.

Purification of recombinant proteins

All proteins were purified as described previously17. Briefly, 10 g E. coli wet cell mass containing the recombinant cyclase was resuspended in 75 mL lysis buffer (10 mM HEPES pH 7.6, 50 mM NaCl, 10% glycerol). Cells were lysed by the addition of lysozyme (0.5 mg/mL) and sonication, and clarified by centrifugation at 60,000 × g for 25 min. Clarified lysate was loaded by gravity onto 8 mL Ni-NTA Superflow resin (Qiagen). The column was washed with 100 mL lysis buffer containing 20 mM imidazole and 50 mL lysis buffer containing 40 mM imidazole. The proteins were eluted with elution buffer (250 mM imidazole, pH 7.9 and 10% glycerol). Fractions containing the purified cyclase were concentrated using Amicon Ultra 15 centrifugal filters and desalted using PD-10 columns (GE Healthcare) equilibrated with storage buffer (10 mM HEPES pH 7.6, 50 mM NaCl). The purified cyclases were drop-frozen in 30-μL aliquots directly into liquid N2 and stored at −80 °C.

Crystallization of SeMet HpiC1 W73M/K132M

Single, diffraction quality crystals of the HpiC1 W73M/K132M selenomethionine derivative were grown in Intelli-Plate 96-2 shallow-well plates (Hampton research) at 20 °C by mixing 1 μL of 11 mg/mL SeMet HpiC1 in storage buffer with 1 μL of a well solution containing 23% PEG 3350, 200 mM CaCl2, 5% trehalose. Sitting droplets were nucleated after 18 h from an earlier spontaneous crystallization event using a cat whisker. Single, rod-shaped crystals grew to approximate dimensions of 50 × 50 × 250 μm after 14 d. 8 μL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 23% PEG 3350, 200 mM $CaCl_2$, and 9.1% trehalose was added directly to the sitting drops, and the crystals were harvested using nylon loops

and vitrified by rapid plunging into liquid nitrogen. SeMet HpiC1 was crystallized in form 1, space group $P2_12_12_1$ with unit cell dimensions of a = 44.9 Å, b = 81.1 Å, c = 131.7 Å, and two chains in the asymmetric unit.

Crystallization of Native HpiC1 (P42)

Single, diffraction quality crystals of native HpiC1 were grown in Intelli-Plate 96-2 shallow-well plates (Hampton research) at 20 °C by mixing 1 μL of 20 mg/mL HpiC1 in storage buffer with 1 μL of a well solution containing 22% PEG 4000, 200 mM CaCl2, 100 mM Tris pH 8.5, and 5% ethylene glycol. Sitting droplets were nucleated after 4 h from an earlier spontaneous crystallization event using a cat whisker. Single, rod-shaped crystals grew to approximate dimensions of 50 × 50 × 150 μm after 7 d. 8 μL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 22% PEG 4000, 200 mM $CaCl_2$, 100 mM Tris pH 8.5, and 15% ethylene glycol was added directly to the sitting drops and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. In these conditions, HpiC1 native crystallized in Form 2, space group P42 with unit cell dimensions of a = 71.3 Å, b = 71.3 Å, c = 80.6 Å, and two chains in the asymmetric unit.

Crystallization of Native HpiC1 (C2)

Single, diffraction quality crystals of native HpiC1 were grown in Intelli-Plate 96-2 shallow-well plates (Hampton research) at 20 °C by mixing 1 μL of 20 mg/mL HpiC1 in storage buffer and 5% DMSO with 1 μL of a well solution containing 22% PEG 4000, 150 mM CaCl2, 100 mM Tris pH 8.5, and 5% ethylene glycol. Sitting droplets were nucleated after 4 h from an earlier spontaneous crystallization event using a cat whisker. Single, diamond-shaped crystals grew to approximate dimensions of 200 × 200 × 100 μm after 7 d. 8 μL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 22% PEG 4000, 150 mM $CaCl_2$, 100 mM Tris

pH 8.5, 15% ethylene glycol, 5% DMSO was added directly to the sitting drops, and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. In these conditions, HpiC1 native was crystallized in form 3, space group C2 with unit cell dimensions of a = 113.8 Å, b = 49.5 Å, c = 53.1 Å, α = 90°, β = 110.5°, γ = 90° and one chain in the asymmetric unit.

Crystallization of HpiC1 Y101F

Single, diffraction quality crystals of HpiC1 Y101F were grown in Intelli-Plate 96-2 shallow-well plates (Hampton research) at 20 °C by mixing 1 μL of 15 mg/mL protein in storage buffer with 1 μL of a well solution containing 22% PEG 4000, 150 mM CaCl2, 100 mM Tris pH 8.5, and 5% ethylene glycol. Sitting droplets were nucleated after 4 h from an earlier spontaneous crystallization event using a cat whisker. Single, diamond-shaped crystals grew to approximate dimensions of 250 × 250 × 270 μm after 7 d. 8 μL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 22% PEG 4000, 150 mM $CaCl_2$, 100 mM Tris pH 8.5, and 15% ethylene glycol was added directly to the sitting drops, and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. HpiC1 Y101F crystallized in form 3, space group C2 with unit cell dimensions of a = 113.8 Å, b = 49.8 Å, c = 53.4 Å, α = 90°, β = 110.4°, γ = 90° and one chain in the asymmetric unit.

Crystallization of HpiC1 Y101S

Single, diffraction quality crystals of HpiC1 Y101S were grown in Intelli-Plate 96-2 shallow-well plates (Hampton research) at 20 °C by mixing 1 μL of 15 mg/mL protein in storage buffer with 1 μL of a well solution containing 20% MEPEG 5000, 150 mM $CaCl_2$, 100 mM Tris pH 8.5, 5% ethylene glycol. Sitting droplets were nucleated after 4 h from an earlier spontaneous crystallization event using a cat whisker. Single, diamond-shaped crystals grew to approximate

dimensions of 250 × 250 × 270 μm after 7 d. 8 μL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 20% MEPEG 5000, 150 mM $CaCl_2$, 100 mM Tris pH 8.5, 15% ethylene glycol, 5% DMSO was added directly to the sitting drops and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. HpiC1 Y101S crystallized in form 3, space group C2 with unit cell dimensions of a = 113.9 Å, b = 49.6 Å, c = 53.4 Å, α = 90°, β = 110.3°, γ = 90° and one chain in the asymmetric unit.

Crystallization of HpiC1 F138S and Y101F/F138S

Single, diffraction quality crystals of HpiC1 F138S and Y101F/F138S were grown in Intelli-Plate 96-2 shallow-well plates (Hampton research) at 20 °C by mixing 1 μL of 15 mg/mL protein in storage buffer, 20 mM $CaCl_2$, 5% DMSO with 1 μL of a well solution containing 20% MEPEG 5000, 100 mM Bis–Tris pH 6.5, and 5% ethylene glycol. Sitting droplets were nucleated after 4 h from an earlier spontaneous crystallization event using a cat whisker. Single, plate-shaped crystals grew to approximate dimensions of 50 × 50 × 300 μm after 7 d. 8 μL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 20% MEPEG 5000, 20 mM $CaCl_2$, 100 mM Bis–Tris pH 6.5, and 15% ethylene glycol, 5% DMSO was added directly to the sitting drops and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. HpiC1 F138S and HpiC1 Y101F/F138S crystallized in form 4, space group P21 with unit cell dimensions of a = 62.0 Å, b = 47.9 Å, c = 174.2 Å, α = 90°, β = 97.2°, γ = 90° and four chains in the asymmetric unit.

Data collection and processing

X-ray data were collected at 100 K on beamline 23ID-B at the General Medical Sciences and Cancer Institute Structural Biology Facility at the Advanced Photon Source in Argonne, IL,

USA. Diffraction data were integrated and scaled using XDS32. Data collection statistics are given in Supplementary Table 1.

Experimental phasing (SAD) and molecular replacement, model building and refinement

The structure of SeMet HpiC1 W73M/K132M was solved using single-wavelength anomalous diffraction (SAD). Phasing and initial model building were performed using Phenix Autosol[33]. This resulted in an initial model that could be extended by alternating cycles of manual building in Coot34 and least-squares refinement with Refmac[35]. The structures for HpiC1 native and Y101F, Y101S, F138S, Y101F/F138S were solved by molecular replacement using Phaser-MR[36] with the structure of the HpiC1 SeMet derivative as a search model. Final models were generated by alternating cycles of manual building in Coot[34] and refinement in Refmac[35] and Phenix[33], and were validated using MolProbity[37].

Docking 12-epi-hapalindole U with Autodock VINA

12-epi-hapalindole U was docked into the SeMet HpiC1 model using Autodock VINA23. Default parameters for Autodock VINA were used, with the exception of exhaustiveness, which was set to 100.

Molecular graphics

All figures depicting the protein structure were generated with PyMOL (Version 1.8 Schrödinger, LLC).

Chemical synthesis

Indole isonitrile was synthesized as described previously[11].

In vitro cyclase assays

In vitro assays were performed with HpiC1 and FimC5 and their corresponding active site mutants as described previously[17]. Products were analyzed using LC–MS (Shimadzu) using C18 (Agilent) HPLC column and monitored by UV absorbance at 280 nm.

Scaleup, purification, and NMR of F138S product (12-epi-fischerindole U)

The semi-prep scale reaction was performed as described previously[17].

Quantum mechanical calculations

Conformational searches of the hapalindole and fischerindole products were performed using the Schrödinger MacroModel software package (release 2017-2, Schrödinger, LLC), and the lowest energy conformation was used for all reported quantum mechanical calculations. All quantum mechanical calculations were performed using the Gaussian 09 (Revision A.02, Gaussian, Inc.) software package. Structures were optimized in the gas phase at the B3LYP38,39/6-31 G(d) level of theory; frequency calculations were used to confirm the presence of local minima (no imaginary frequencies) and transition states (one imaginary frequency) and to calculate free energies at 298 K. To obtain more accurate energetics, single-point energy calculations were performed on the optimized structures at the B3LYP/6-311 + + G(d,p) level of theory using Grimme's D3(BJ) dispersion correction40,41 and the IEFPCM[42,43,44] solvent model for diethyl ether ($\varepsilon = 4$). The use of the dielectric constant $\varepsilon = 4$ has proven to be a good model to estimate the dielectric permittivity in the enzyme active site, accounting for electronic polarization and small backbone fluctuations[45,46].

Molecular dynamics simulations

Molecular dynamics simulations were performed using the GPU code (pmemd)[47] of the AMBER 16 package (AMBER 2016, University of California). Parameters for intermediates and substrates were generated within the antechamber module using the general AMBER force field

(gaff)[48], with partial charges set to fit the electrostatic potential generated at the HF/6-31 G(d) level by the RESP model49. The charges were calculated according to the Merz–Singh–Kollman scheme[50,51] using the Gaussian 09 package (Revision A.02, Gaussian, Inc.). Each protein was immersed in a pre-equilibrated truncated cuboid box with a 10 Å buffer of TIP3P52 water molecules using the leap module, resulting in the addition of around 15,000 solvent molecules. The systems were neutralized by addition of explicit counterions (Na^+ and Cl^-). All subsequent calculations were done using the widely tested Stony Brook modification of the Amber14 force field (ff14sb)53. A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on the solute by a harmonic potential with a force constant of 500 kcal·mol−1·Å−2 and the second stage minimizes all the atoms in the simulation cell. The systems were gently heated using six 50 ps steps, incrementing the temperature by 50 K for each step (0–300 K) under constant-volume and periodic-boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between the hydrogen atoms was kept fixed. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method54. An 8 Å cutoff was applied to Lennard–Jones and electrostatic interactions. Harmonic restraints of 10 kcal·mol–1 were applied to the solute and the Langevin equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated without restrains for 2 ns with a 2-fs time step at a constant pressure of 1 atm and temperature of 300 K. After the systems were equilibrated in the NPT ensemble, subsequent MD simulations were performed for an additional 500 ns under an NVT ensemble and periodic-boundary conditions.

3.4 Results and Discussion

Following recent work reporting the function and selectivity of Stig cyclase proteins[11,17,18,19], we sought a structure of HpiC1 to understand the basis for the complex cyclization cascade of 1 and to comprehend how the Stig cyclases catalyze formation of various alkaloid products from this common biosynthetic intermediate. We obtained HpiC1 crystals in four different forms under conditions with varied Ca^{2+} concentrations. These forms reveal the same overall structure for HpiC1, but differ in their space group, indicating that Ca^{2+} concentration can influence crystal packing (see below). The 1.7-Å structure in form 1 was solved by selenomethionyl (SeMet) SAD phasing from a HpiC1 W73M/K132M double mutant, as the wild-type protein lacks Met. The SeMet W73M/K132M structure was used to solve structures in the other crystal forms by molecular replacement. The overall fold of the HpiC1 polypeptide is a flattened β-jelly roll fold (Fig. 2a) composed of two antiparallel β-sheets. The antiparallel pairing of the β6 strands from two monomers creates a continuous β-sheet across an extensive dimer interface, which buries 2,060 $Å^2$ of total surface area (PISA, Fig. 2b)[20] and encompasses approximately 20% of the total surface area of each monomer. This arrangement is consistent with that seen using size-exclusion chromatography, wherein HpiC1 and other Stig cyclases migrated as apparent dimers[17]. HpiC1 shares the highest structural similarity (2.3 Å r.m.s. deviation from DALI[21]) with the carbohydrate-binding module (CBM) from xylanase in the thermostable bacterium *Rhodothermus marinus* (PDB ID 2Y64; Fig. 2c)[22]. These proteins have highly similar tertiary structure and topology, with the most substantial differences occurring at their N termini and in the loop regions between shared β-strands.

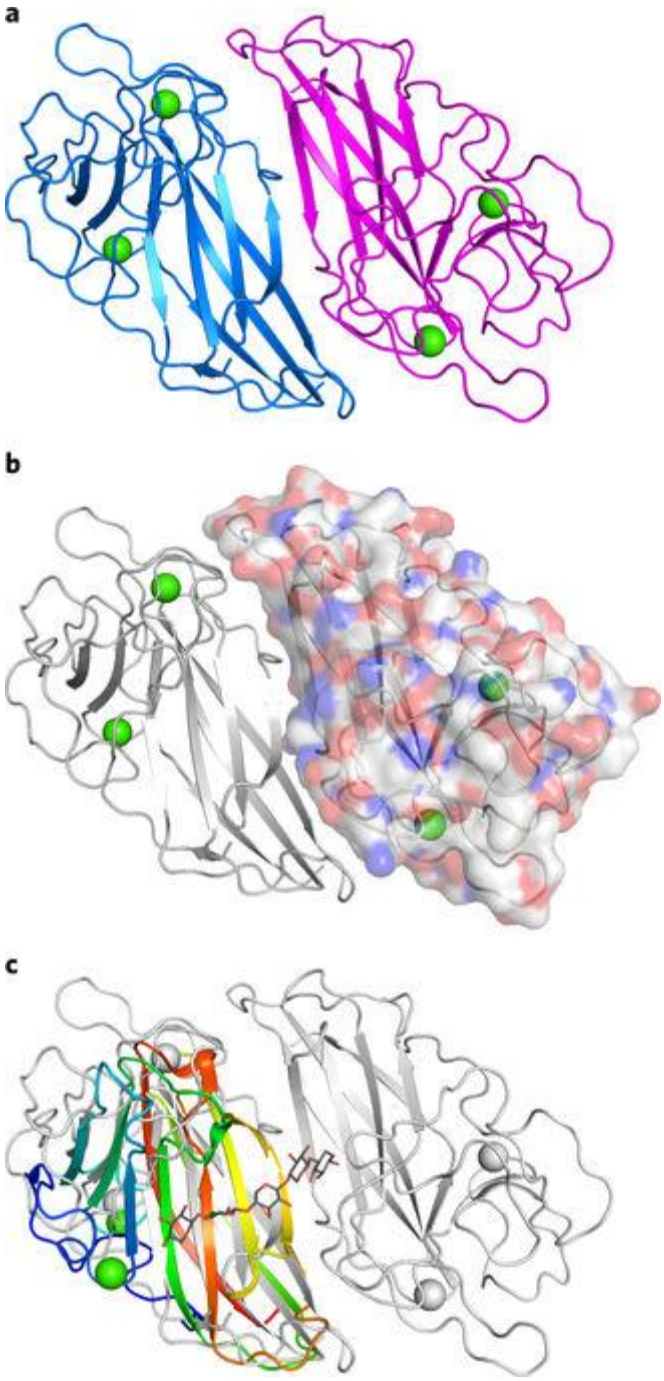
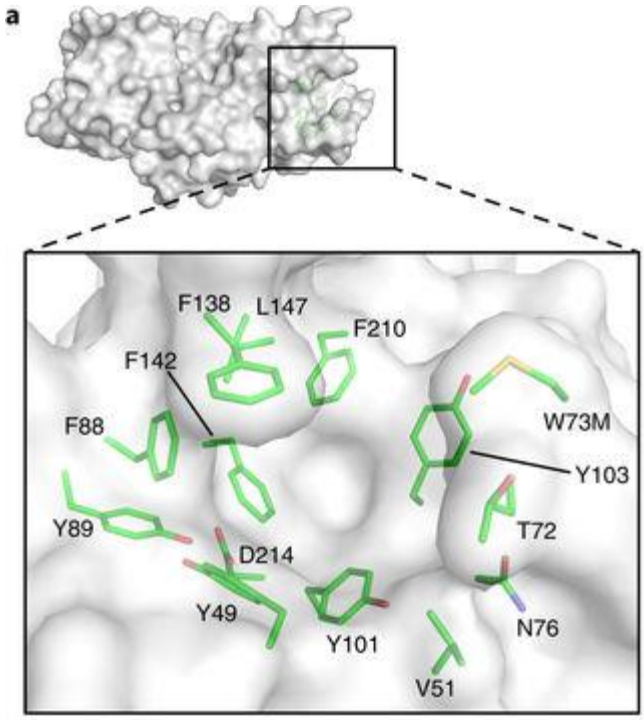


Figure 3.2. a, Cartoon representation of the HpiC1 homodimer. The subunits are colored blue and magenta; green spheres indicate bound calcium ions. b, Surface representation of a single HpiC1 protomer, colored by atom, indicates 2,060 Å2 of buried surface area between the subunits. c, Superposition of HpiC1 with a xylanase carbohydrate-binding module (CBM) homolog (PDB ID 2Y64, rainbow); CBM is monomeric despite sharing the same fold as HpiC1.

HpiC1 has two integral Ca^{2+} ions, each with octahedral coordination geometry. These sites were a key starting point for assessing the structural and catalytic role that Ca^{2+} plays in the Stig cyclase enzymes. In vitro assays conducted in the presence of 5 mM EDTA showed no activity, confirming that Ca^{2+} is required for catalytic function in HpiC1[18]. From sequence comparisons, we expect the integral calcium-binding sites and the core dimeric assembly to be maintained in the other Stig cyclase enzymes.

The substrate-free HpiC1 structure did not immediately suggest the location(s) of substrate binding and catalysis. Initial localization of the active site was revealed by abnormal electron density in a pocket located at the distal end of each subunit. Polyethylene glycol from the crystallization was modeled into this pocket, composed of numerous aromatic amino acids (Fig. 3). Given the hydrophobicity of substrate 1, we explored whether this hydrophobic pocket could be the active site. This region was probed using AutoDock VINA with 5 (Fig. 1)[23], and the major product was chosen based on its defined stereochemistry and rigid scaffold. The top docking solutions had affinities ranging between –9.7 and –9.4 kcal/mol, suggesting that this site possesses an appropriate size and shape to accommodate the hapalindole core.



b

	Y49	V51	T72	W73	N76	F88	Y89	Y101	Y103	F138	F142	L147	F210	D214
HpiC1	Y	V	TW	N	FY	Y	Y	F	F	L	F	D
FamC1	Y	I	TW	N	FY	Y	Y	L	F	F	F	D
FilC1	Y	V	TW	N	FY	Y	Y	L	F	F	F	D
FimC1	Y	I	TW	N	FY	Y	Y	L	F	F	F	D
HpiC2	Y	I	KW	N	FY	F	Y	F	F	F	F	D
FamC2	Y	I	KW	N	FY	F	Y	F	F	F	F	D
FilC2	Y	I	KW	N	FY	F	Y	F	F	F	F	D
FimC2	Y	I	KW	N	FY	F	Y	F	F	F	F	D
HpiC3	Y	V	VQ	Y	FY	S	Y	S	Y	L	S	D
FamC3	Y	V	VQ	Y	FY	S	Y	S	Y	L	S	D
FilC3	Y	V	VQ	Y	YY	S	Y	S	Y	L	S	D
FimC3	Y	V	VQ	Y	FY	S	Y	-	-	-	-	-
HpiC4	F	L	DS	Y	FY	S	Y	A	F	L	S	D
FamC4	F	L	NS	Y	FY	S	Y	A	F	L	S	D
FilC4	F	L	DS	Y	FY	S	Y	A	F	L	S	D
FimC4	F	L	DS	Y	FY	S	Y	A	F	L	S	D
HpiC5	Y	I	RI	N	FY	F	Y	S	F	L	F	D
FimC5	Y	I	RI	N	FY	F	Y	S	F	L	F	D

Figure 3.3. a, Surface representation of the SeMet HpiC1 active site. Key residues are shown as green sticks. Met73 is substituted for the native tryptophan residue. This mutant protein retained wild-type activity. b, Key active site residues shown in an alignment with those from other Stig cyclases. Residues are colored by conservation and side chain composition (ClustalX). The FimC3 gene product is truncated (dashes).

The docking solutions compelled us to further interrogate this region as the putative active site. First, we probed the catalytic role of Asp214 in HpiC1. This residue is 100% conserved in all the currently identified Stig cyclases and is exceptional in that it is the amino acid in the hydrophobic pocket most likely to participate in acid–base chemistry, which can be inferred to promote [3,3]-sigmatropic rearrangments[24]. The Asp214 carboxylate lacks a counter ion and is hydrogen bonded to the Tyr89 hydroxyl. Substitution of Asp214 to alanine abolished activity (Fig. 4), indicating that it plays a critical role in the catalytic cascade. The corresponding Asn214 and Glu214 variants were also completely inactive (Supplementary Fig. 6). These results provide compelling evidence that this hydrophobic pocket is the enzyme active site.

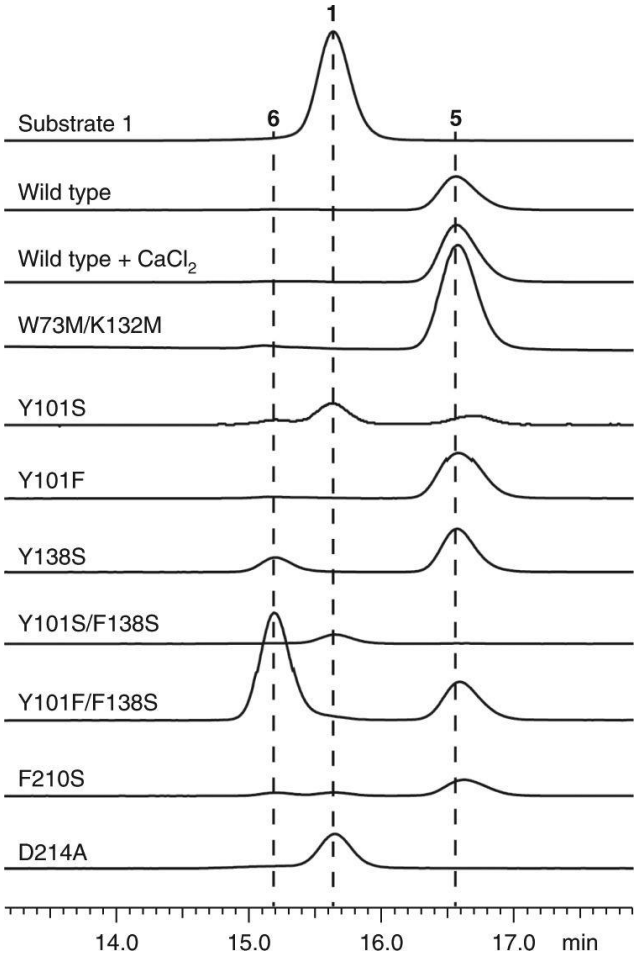


Figure 3.4. Substitution of the catalytic acid Asp214 with alanine abolished activity. A single mutation, F138S, altered the native product profile of HpiC1 to produce 6, which became

predominant in the HpiC1 Y101F/F138S double mutant. Products were monitored by HPLC and confirmed using 1H NMR.

The enzyme conformation around Asp214 varies among the crystal forms. In form 1, Asp214 is shielded by Phe138, whereas in forms 2–4 this residue is shifted to expose Asp214 to the binding pocket. The loop containing Phe138 is also shifted ~3 Å from its position in form 1, although coordination of the calcium ion by the carbonyl oxygen of Phe138 is maintained. The atomic temperature factors in this region also indicate a higher degree of mobility (Supplementary Fig. 8). It is apparent that conformational flexibility is required to perform multiple catalytic steps in a single binding site and to accommodate the broad range of regio- and stereochemical configurations in this class of alkaloids[1].

Based on the different product profiles of the various Stig cyclases, we examined the conservation of residues in this hydrophobic pocket. A comparison revealed that sequence clustering correlates to similar product profiles. We reasoned that key determinants of cyclase reactivity were contained within these localized sequences (Fig. 3), and these residues were explored by mutagenesis. A comparative analysis between HpiC1 and FimC5 (68% sequence identity with HpiC1; catalyzes production of 6 from 1) was pursued on the basis of their ability to differentially produce hapalindole or fischerindole core ring systems, respectively (Fig. 1)[17]. Remarkably, substitution of Phe138 in HpiC1 to the corresponding serine from FimC5 led to the generation of a mixture of its major product, hapalindole 5, and the FimC5 major product, fischerindole 6 (Fig. 4). The product ratio in F138S is approximately 1:2 (6:5) and is shifted to 2:1 (6:5) in Y101F/F138S. HpiC1 Y101F had a product profile comparable to that of the wild-type protein, indicating that this effect is driven primarily by Phe138. A corresponding mutation at this position in FimC5, S139F, did not lead to formation of 5; instead, the product profile was shifted toward the production of 7, a minor tricyclic shunt product of the native HpiC1 and FimC5

reactions (Supplementary Fig. 9)[17]. These data indicate that HpiC1 Phe138 and FimC5 Ser139 both play a key role in directing terminal electrophilic aromatic substitution.

Mutations guided by sequence alignments were introduced into HpiC1 to identify additional key residues (Fig. 3). The Y101S/F138S variant corresponding to the FamC3, HpiC3, and FilC3 homologs showed no activity with 1 (Fig. 4), which is consistent with the lack of reactivity observed in homodimeric forms of FamC3, HpiC3, and FilC3 [17]. HpiC1 Y101S also showed reduced activity, though to a lesser extent than in combination with F138S. Intriguingly, FamC3 was shown to associate with FamC2 as a heterodimer, and catalyzes the formation of hapalindole H[17].

To understand the impact of these mutations, high-resolution structures were determined for HpiC1 variants Y101F, Y101S, F138S, and Y101F/F138S. Whereas our efforts to observe bound ligands by either soaking or co-crystallization only afforded complexes with DMSO and Tris buffer, these structures aided our efforts to interrogate the mechanism of cyclization in HpiC1 using computational methods.

We first applied molecular dynamics (MD) simulations to the substrate-free structures to gain insights into the dynamics of the active site. Starting from the apo-HpiC1 dimeric structure, analysis of the MD trajectories revealed large fluctuations of the loop containing Phe138 (Asn137–Phe150), which is in agreement with the different conformations found for this loop in the crystal structures. MD simulations showed that Asp214, which is essential for enzyme activity, stays preferentially in a conformation in which the Asp214 side chain points toward the inner cavity of the active site, while Phe138 acts as a wall on the side of the active site pocket. This is due, in part, to the Tyr89 hydroxyl hydrogen bond with the Asp214 carboxylic acid group. Importantly, the predicted pK_a for Asp214, estimated from snapshots obtained along 500 ns of MD simulation, is

6.5–7.0, indicating that it can be protonated in an acid–base equilibrium to act as a protonating species during catalysis. An alternative conformation of Phe138 is sampled during the 500-ns trajectory. In this conformation, the Phe138 side chain is displaced and Asp214 becomes inaccessible, generating an inactive conformation that is similar to the arrangement observed in the crystal form 1.

We next considered the lack of activity in the Y101S/F138S mutant. MD simulations for Y101S/F138S showed that Y101S interacts closely with Asp214, as well as with the Tyr89 hydroxyl hydrogen bond. These two hydrogen bonds favor stabilization of a negatively charged carboxylate group, which would not be protonated in this more polar environment. This is confirmed by the decrease of the predicted pK_a value of Asp214 (pK_a ≈5.5–6).

As described above, F138S and Y101F/F138S mutants change the product profile in HpiC1, leading to increased formation of fischerindole 6 (Fig. 4). We performed 500-ns MD simulations on both F138S and Y101F/F138S mutants and found critical changes in the shape of the active site. The F138S mutation creates more space around the catalytic Asp214 residue and releases the interaction between the two phenyl rings of Phe138 and Phe210, which results in Phe210 becoming more flexible. This active site reshaping is responsible for the change in the reaction outcome, as discussed below.

We employed DFT calculations to explore the possible reaction mechanism for Stig cyclases, in particular, the Cope rearrangement, which is the first step in the three-part reaction cascade starting from 1 [11]. The instability of 1 has precluded determination of its chiral configuration at the indolenine C3 position. Thus, we computed the Cope rearrangement and cyclization cascade mechanism (Fig. 5) starting from both the (*R*)-1 and (*S*)-1 enantiomers.

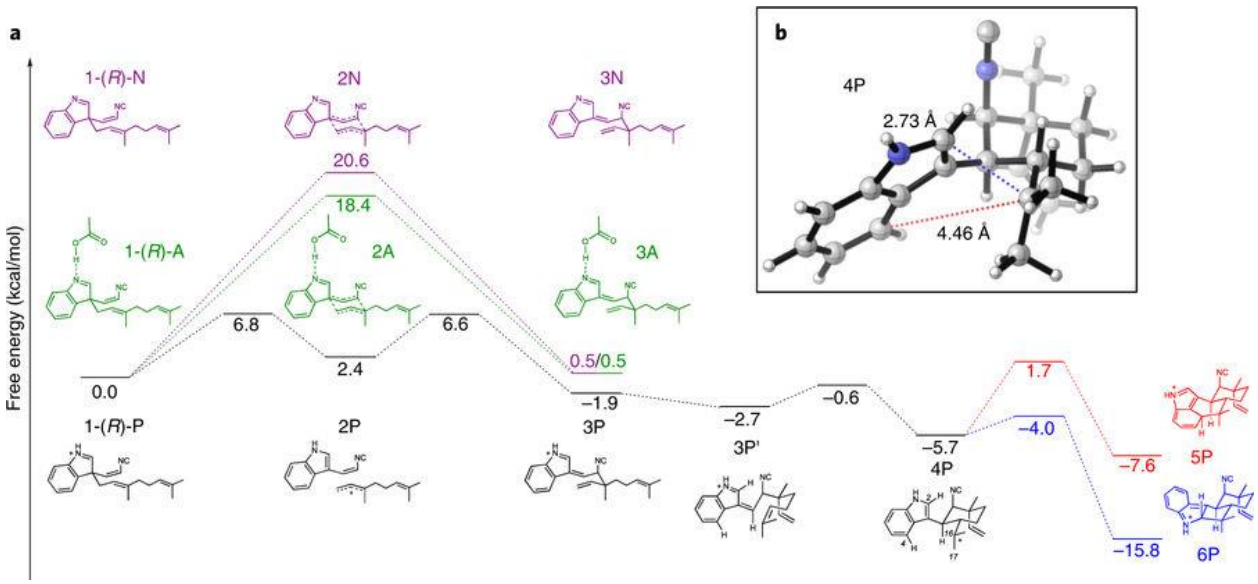


Figure 3.5. a, Cope rearrangement, 6-exo-trig cyclization, and electrophilic aromatic substitution cascade starting from the R enantiomer of substrate 1 in a near-attack conformation, leading to 5 precursor 5P and 6 precursor 6P. The energetics of the Cope rearrangement are computed with the neutral indolenine (pathway N), the N-protonated indolenine (pathway P), and the indolenine forming a hydrogen bond with acetic acid (pathway A). b, Optimized geometry of key intermediate 4P, which undergoes regioselective electrophilic aromatic substitution to form 5 or 6.

Although a typical Cope rearrangement has four possible transition states, two chairs and two boats, only one chair-like transition state can account for the known stereochemistry at C11 and C12 in the products, and therefore only this transition state was further considered. Given that the active site of HpiC1 contains the essential Asp214 residue, we explored the impact of this residue on accelerating the Cope rearrangement. The conversion of neutral starting material 1-(*R*)-N to intermediate 3N is concerted, proceeding through a single chair-like transition state, 2N, which lies 20.6 kcal/mol above the near-attack conformation of the starting material. This chair-like transition structure has dissociative character, breaking and forming partial single bonds of 2.58 Å and 2.53 Å, respectively. The partial negative charge on the isocyanidovinylindolenine fragment of −0.43 *e* can be stabilized by a hydrogen bonding donor at the indolenine nitrogen, which has a partial negative charge of –0.49 *e*. Adding an acetic acid molecule to mimic possible hydrogen bonding between Asp214 and the indolenine nitrogen lowers this barrier by 2.2 kcal/mol

(a 40-fold rate enhancement), with the conversion of 1-(*R*)-A to 3 A proceeding through a chair-like transition state, 2A, with a free-energy barrier of 18.4 kcal/mol. These computations suggest that Asp214 can facilitate the Cope rearrangement by hydrogen bonding to the indolenine nitrogen. Fully protonating the indolenine nitrogen, which represents the maximum limit of potential acid catalysis by Asp214, results in a change of mechanism. The conversion from 1-(*R*)-P to 3P is stepwise and dissociative, with 2P being an intermediate rather than a transition state, and has a much lower overall free-energy barrier of 6.8 kcal/mol. Intermediate 2P is stabilized by full conjugation between the indole and isonitrile groups, as well as by an allylic cation to produce this much lower overall barrier (Fig. 5).

The second proposed step in the biosynthesis of 5 and 6 is the 6-*exo*-*trig* cyclization of intermediate 3 to 4, which sets the stereochemistry at C10 and C15 in the products. In gas-phase DFT optimizations, it was possible to locate a transition state for this cyclization only when the indolenine nitrogen was protonated. Without protonation, the zwitterionic character of the possible transition state leads to bond formation between negatively charged position C3 of the indolenine and positively charged position C16 to generate a cyclobutane ring. By contrast, the protonated species undergoes facile cyclization from 3P to intermediate 4P through a low-lying transition state. This suggests that protonation is crucial and that Asp214 may catalyze cyclization, and possibly the preceding Cope rearrangement, in this way.

The third step is electrophilic aromatic substitution of intermediate 4P, whereby two different transition states lead to the two major products. Electrophilic aromatic substitution at C4 of the indole yields intermediate 5P, which gives 5 upon deprotonation, whereas electrophilic aromatic substitution at C2 of the indole yields intermediate 6P, which gives 6 upon deprotonation. Deprotonation at C16 can lead to formation of tricyclic 7 (Fig. 1), which is generated as a trace

product by HpiC1. Quantum mechanics calculations show that formation of 6, which has the lower energy transition state, should be intrinsically favored. Thus, the regioselectivity of electrophilic aromatic substitution to generate 5 appears to be controlled by the HpiC1 active site as opposed to inherent energetics of the system.

To understand enzymatic control of the Cope rearrangement, we conducted 500-ns MD simulations on wild-type HpiC1 with both 1-(*R*)-P and 1-(*S*)-P bound into the active site. Both enantiomers retain the key hydrogen bond interaction between the protonated indolenine NH and the Asp214 residue. However, the C11–C12 distance, which corresponds to the C–C bond formed during the Cope rearrangement, is shorter (~3.5 Å) for the 1-(*R*)-P substrate than for the 1-(*S*)-P substrate (>4.0 Å). In addition, only 1-(*R*)-P is stabilized by the enzyme active site in a near-attack conformation that leads to the correct stereochemistry at positions 11 and 12 (Fig. 6a,b). Based on these observations, the *R* enantiomer of 1 is the most plausible natural substrate.

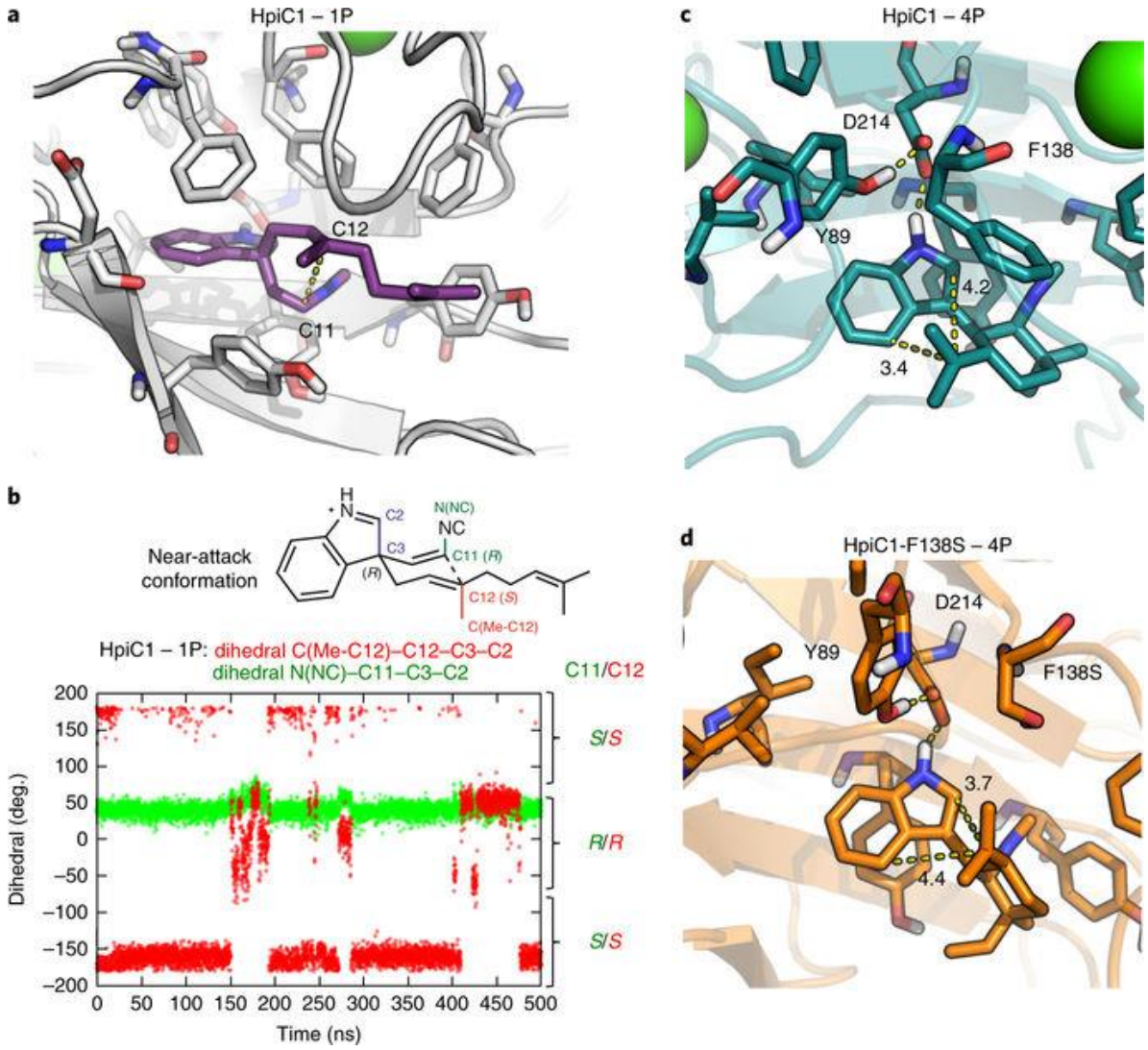


Figure 3.6. a,b, Representative snapshots of substrate 1-(R)-P (violet) bound into HpiC1 (a) and dihedral angles explored during the 500 ns of MD simulation for substrate 1-(R)-P bound into HpiC1 (b). Dihedral-1 (N(NC)–C11–C3–C2) and dihedral-2 (C(Me-C12)–C12–C3–C2) define the relative orientation of substituents at the C11 and C12 positions, respectively, during the MD simulation. The right axis indicates the final stereochemistry of C11 and C12 expected after the Cope rearrangement coming from the given near-attack conformation of 1-(R)-P, as shown in b and Supplementary Fig. 22. 1-(R)-P mainly explores one conformation during the MD trajectory, leading to R configuration at C11 and S configuration at C12. c,d, Representative snapshots obtained from 500 ns of MD simulations of the active site for intermediate 4P bound to HpiC1 (c; Supplementary Figs. 23 and 24) and F138S (d; Supplementary Fig. 26). Distances (in Ångstroms) show that the conformation adopted by intermediate 4P in the HpiC1 wild-type enzyme moves C16 closer to C4 to form the hapalindole product, while the F138S mutant enables the exploration of a conformation of intermediate 4P in which C16 gets closer to C2 to allow fischerindole product formation.

To understand how HpiC1 could control the differentiation of hapalindole and fischerindole products, we performed MD simulations with intermediates before electrophilic aromatic substitution. We considered the two intermediate precursors 4P and 10P, which derive from the 1-(*R*)-P and 1-(*S*)-P starting materials, respectively. As 4P or 10P can each generate both the 5 and 6 products, we analyzed the binding of both intermediates. MD simulations with the two docked intermediates show that 4P effectively maintains the hydrogen bond between the protonated NH-indole and Asp214 more than 10P, which rapidly dissociates during the simulation. These results reinforce the idea that (*R*)-1 is the most plausible natural substrate. Moreover, MD simulations showed that when 4P is bound into the active site, it adopts a conformation in which the distance between C4 and C16 (~3.5 Å; corresponds to hapalindole formation) is shorter than the distance between C2 and C16 (~4.0 Å; corresponds to fischerindole formation) (Fig. 6c). Finally, at around 200 ns, a conformational change of the Phe138 side chain reverses this trend to disrupt the interaction of 4P and Asp214, highlighting the key role of Phe138 in controlling the site selectivity of the reaction.

The critical role of Phe138 was investigated through MD simulations on the F138S and Y101F/F138S mutants using 4P. In both trajectories, the distance between C2 and C16 leading to fischerindole formation is shortened compared to that in the wild-type enzyme, becoming closer (~3.5 Å) than the distance between C4 and C16 (~4.0 Å). This leads to hapalindole formation, consistent with the experimentally observed product ratios (Fig. 6d). The absence of the bulky Phe138 residue near C2 allows the intermediate to adopt a slightly different conformation, enabling formation of 6 without completely suppressing generation of 5 (Fig. 6d).

As with the terpene cyclases[25,26,27], aromatic side chains in HpiC1 are appropriately positioned to form cation–π interactions with the carbocation of 4P. Examination of the MD

trajectories shows that Phe138 and Phe142 can interact with the C16 carbocation, whereas Phe88 and Phe210 are either too far away or poorly oriented for interaction[28]. An aromatic amino acid Phe142 is conserved in all known Stig cyclases, whereas Phe138 varies with different product outcomes, as observed in our mutational analysis and calculations.

3.5 Discussion

The structure of HpiC1 has provided the first high-resolution insights into a fascinating mechanistic puzzle in which the Stig cyclases are able to generate extensive stereochemical and regiochemical diversity through the common biosynthetic precursor 1. The surprising function of the Stig cyclases[11] was not able to be inferred from bioinformatic analysis, and, similarly, structural and mechanistic insights from homology-based tools were unavailable for these remarkable biocatalysts. HpiC1 is not homologous to any characterized terpene cyclase, but instead is most similar to bacterial CBMs, which function primarily to bring various hydrolytic enzymes into contact with their carbohydrate substrates. Target sugar molecules bind the CBMs in an extended cleft at the protein surface, and this is mediated through several amino acids that are not conserved in HpiC1 (Fig. 2c)[22]. This indicates a divergent functionality in HpiC1 based on a shared protein scaffold. Nevertheless, Stig cyclases and CBMs share several properties including a common fold, thermostability[29], and structural calcium ions, which play a key role in CBM stabilization, substrate recognition, and oligomerization[29,30].

The molecular basis for the Ca^{2+} dependence of HpiC1 was demonstrated through two well-ordered binding sites near the enzyme active site. A paradoxical aspect of this calcium requirement is the observation that low millimolar concentrations of $CaCl_2$ caused HpiC1 to precipitate. This effect was reversible through stoichiometric addition of EDTA, indicative of calcium-dependent higher order oligomerization of HpiC1. However, a requirement for millimolar

concentrations of Ca^{2+} in the reaction to achieve activity has been reported for some cyclases[18]. There are also fascinating instances of heteromeric association of some Stig cyclases leading to variant stereochemical outcomes of the products compared to the products of their homomeric counterparts[17,18]. Taken together, these data are indicative of an important, yet complex, structural role for Ca^{2+} in Stig cyclase catalysis that may involve higher order oligomerization. Because HpiC1 could not be crystallized without supplemental $CaCl_2$, we examined the lattice contacts for evidence of additional calcium binding sites. In crystal forms 1 and 2, which required 200 mM Ca^{2+} in the well solution, we found two sites in which calcium was able to facilitate formation of higher order complexes. We observed one fully occupied interfacial calcium ion far from the active site in crystal form 1 and a second Ca^{2+} at half occupancy in unrelated crystal forms 1 and 2. This Ca^{2+} is adjacent to the active sites, indicating that a bridging calcium ion could influence the activity of cyclase oligomers. A functional role of higher order oligomerization in HpiC1 remains unclear, as the addition of supplemental (1–20 mM) Ca^{2+} enhances the enzymatic activity of HpiC1 but is not required for turnover (Fig. 4). Further structural investigation will be required to understand the assembly of heteromeric Stig cyclase complexes and the relevance of Ca^{2+} in those cases.

We have established the location of the HpiC1 active site and, using mutational analysis, identified critical residues for catalysis and demonstrated a key relationship between amino acid sequence and product outcome. Most importantly, we identified Asp214 as the source of an active site acid required for catalysis and is consistent with reports of acid-catalyzed Cope rearrangements[24]. An active site acid is also consistent with the reported three-fold increase in apparent k_{cat} as pH was decreased from 9.0 to 6.0 in the Stig cyclase WelU1 (68% sequence identity with HpiC1)[19]. Notably, the hydrophobic environment around Asp214 is essential for maintaining

a suitable population of the protonated species, as the Y101S/F138S mutation substantially reduces enzyme activity. We also identified a key regiochemical switch at Phe138 that gave rise to production of the fischerindole core in HpiC1. In these studies, the general importance of Phe138 in the HpiC1 homolog FimC5 was also confirmed, whereby a corresponding mutation also affected the product distribution with respect to regiochemistry. Together, these findings will facilitate our efforts to anticipate the product profiles in new cyclases that are identified as additional strains and gene clusters are discovered.

We also explored the HpiC1 active site dynamics using a combination of DFT quantum mechanical calculations of mechanism and modes of catalysis, and MD simulations on the protein. We studied the origins of the three-part catalytic mechanism, and how HpiC1 controls the regiochemistry of product formation by favoring a particular conformation of substrate 1 and the reaction intermediate 4P. We have examined the role of key mutations in HpiC1 that switch the native product outcome from hapalindole 5 to fischerindole 6. Together, these results address several of the catalytic steps in the formation of 5 from (*R*)-1. Further analysis will be required to establish the mechanistic basis by which the variant Stig cyclases achieve differentiation at the 6-*exo-trig* cyclization step, where the stereocenters at C10 and C15 are set (for example, Hapalindole U, H, and J series). Structural studies on additional Stig cyclases, mutational analysis across key active site residues, and computational modeling of the reaction intermediates will enable prediction of product profiles, and engineering of new selectivities to diversify further this remarkable family of natural products.

3.6 References

1. Bhat, V., Dave, A., MacKay, J. A. & Rawal, V. H. The chemistry of hapalindoles, fischerindoles, ambiguines, and welwitindolinones. Alkaloids Chem. Biol. 73, 65–160 (2014).

2. Asthana, R. K. et al. Identification of an antimicrobial entity from the cyanobacterium Fischerella sp. isolated from bark of Azadirachta indica (Neem) tree. J. Appl. Phycol. 18, 33–39 (2006).

3. Becher, P. G., Keller, S., Jung, G., Süssmuth, R. D. & Jüttner, F. Insecticidal activity of 12-epi-hapalindole J isonitrile. Phytochemistry 68, 2493–2497 (2007).

4. Cagide, E. et al. Hapalindoles from the cyanobacterium Fischerella: potential sodium channel modulators. Chem. Res. Toxicol. 27, 1696–1706 (2014).

5. Mo, S., Krunic, A., Chlipala, G. & Orjala, J. Antimicrobial ambiguine isonitriles from the cyanobacterium Fischerella ambigua. J. Nat. Prod. 72, 894–899 (2009).

6. Mo, S., Krunic, A., Santarsiero, B. D., Franzblau, S. G. & Orjala, J. Hapalindole-related alkaloids from the cultured cyanobacterium Fischerella ambigua. Phytochemistry 71, 2116–2123 (2010).

7. Zhang, X. & Smith, C. D. Microtubule effects of welwistatin, a cyanobacterial indolinone that circumvents multiple drug resistance. Mol. Pharmacol. 49, 288–294 (1996).

8. Hillwig, M. L., Zhu, Q. & Liu, X. Biosynthesis of ambiguine indole alkaloids in cyanobacterium Fischerella ambigua. ACS Chem. Biol. 9, 372–377 (2014).

9. Raveh, A. & Carmeli, S. Antimicrobial ambiguines from the cyanobacterium Fischerella sp. collected in Israel. J. Nat. Prod. 70, 196–201 (2007).

10. Stratmann, K. et al. Welwitindolinones, unusual alkaloids from the blue-green algae Hapalosiphon welwitschii and Westiella intricata. Relationship to fischerindoles and hapalinodoles. J. Am. Chem. Soc. 116, 9935–9942 (1994).

11. Li, S. et al. Hapalindole/ambiguine biogenesis is mediated by a Cope rearrangement, C–C bond-forming cascade. J. Am. Chem. Soc. 137, 15366–15369 (2015).

12. Cope, A. C. & Hardy, E. M., The introduction of substituted vinyl groups. V. A rearrangement involving the migration of an allyl group in a three-carbon system. J. Am. Chem. Soc. 62, 441–444 (1940).

13. Ilardi, E. A., Stivala, C. E. & Zakarian, A. [3,3]-Sigmatropic rearrangements: recent applications in the total synthesis of natural products. Chem. Soc. Rev. 38, 3133–3148 (2009).

14. DeClue, M. S., Baldridge, K. K., Künzler, D. E., Kast, P. & Hilvert, D. Isochorismate pyruvate lyase: a pericyclic reaction mechanism? J. Am. Chem. Soc. 127, 15002–15003 (2005).

15. Luk, L. Y., Qian, Q. & Tanner, M. E. A Cope rearrangement in the reaction catalyzed by dimethylallyltryptophan synthase? J. Am. Chem. Soc. 133, 12342–12345 (2011).

16. Tanner, M. E. Mechanistic studies on the indole prenyltransferases. Nat. Prod. Rep. 32, 88–101 (2015).

17. Li, S. et al. Decoding cyclase-dependent assembly of hapalindole and fischerindole alkaloids. Nat. Chem. Biol. 13, 467–469 (2017).

18. Zhu, Q. & Liu, X. Discovery of a calcium-dependent enzymatic cascade for the selective assembly of hapalindole-type alkaloids: On the biosynthetic origin of hapalindole U. Angew. Chem. Int. Edn Engl. 56, 9062–9066 (2017).

19. Zhu, Q. & Liu, X. Molecular and genetic basis for early stage structural diversifications in hapalindole-type alkaloid biogenesis. Chem. Commun. (Camb.) 53, 2826–2829 (2017).

20. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797 (2007).

21. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, W545–W549 (2010).

22. von Schantz, L. et al. Structural basis for carbohydrate-binding specificity--a comparative assessment of two engineered carbohydrate-binding modules. Glycobiology 22, 948–961 (2012).

23. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461 (2010).

24. Lutz, R. P. Catalysis of the Cope and Claisen rearrangements. Chem. Rev. 84, 205–247 (1984).

25. Wendt, K. U., Poralla, K. & Schulz, G. E. Structure and function of a squalene cyclase. Science 277, 1811–1815 (1997).

26. Starks, C. M., Back, K., Chappell, J. & Noel, J. P. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science 277, 1815–1820 (1997).

27. Lesburg, C. A., Zhai, G., Cane, D. E. & Christianson, D. W. Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. Science 277, 1820–1824 (1997).

28. Jenson, C. & Jorgensen, W. L. Computational investigations of carbenium ion reactions relevant to sterol biosynthesis. J. Am. Chem. Soc. 119, 10846–10854 (1997).

29. Abou-Hachem, M. et al. Calcium binding and thermostability of carbohydrate binding module CBM4-2 of Xyn10A from Rhodothermus marinus. Biochemistry 41, 5720–5729 (2002).

30. Montanier, C. Y. et al. A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization. J. Biol. Chem. 286, 2499–22509 (2011).

31. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. 229, 105–124 (1993).

32. Kabsch, W. XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132 (2010).

33. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 (2010).

34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501 (2010).

35. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. 67, 355–367 (2011).

36. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).

37. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21 (2010).

38. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. 98, 5648–5652 (1993).

39. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B Condens. Matter 37, 785–789 (1988).

40. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 132, 154104 (2010).

41. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. J. Comput. Chem. 32, 1456–1465 (2011).

42. Miertuš, S., Scrocco, E. & Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilizaion of ab initio molecular potentials for the prevision of solvent effects. Chem. Phys. 55, 117–129 (1981).

43. Miertus, S. & Tomasi, J. Approximate evaluations of the electrostatic free-energy and internal energy changes in solution processes. Chem. Phys. 65, 239–245 (1982).

44. Pascual-ahuir, J. L., Silla, E. & Tuñon, I. GEPOL: An improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. J. Comput. Chem. 15, 1127–1138 (1994).

45. Li, L., Li, C., Zhang, Z. & Alexov, E. On the dielectric "constant" of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi. J. Chem. Theory Comput. 9, 2126–2136 (2013).

46. Schutz, C. N. & Warshel, A. What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins 44, 400–417 (2001).

47. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. J. Chem. Theory Comput. 9, 3878–3888 (2013).

48. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. J. Comput. Chem. 25, 1157–1174 (2004).

49. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP model. J. Phys. Chem. 97, 10269–10280 (1993).

50. Besler, B. H., Merz, K. M. & Kollman, P. A. Atomic charges derived from semiempirical methods. J. Comput. Chem. 11, 431–439 (1990).

51. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. J. Comput. Chem. 5, 129–145 (1984).

52. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935 (1983).

53. Maier, J. A. et al. Ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. J. Chem. Theory Comput. 11, 3696–3713 (2015).

54. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092 (1993).

Chapter 4. Mechanism of Stereoselective Catalysis by Diels–Alderase PyrI4 Involved in Pyrroindomycins Biosynthesis

4.1 Abstract

Enzyme PyrI4 catalyzes an intramolecular Diels–Alder (DA) reaction in the biosynthesis of the pyrroindomycin natural products. This enzymatic reaction leads exclusively to the formation of one of the four possible spirocyclic products. We have studied the reaction mechanism of catalysis and origins of stereoselectivity with density functional theory (DFT) calculations and molecular dynamics (MD). MD simulations reveal that PyrI4 binds preferentially the reactive conformation that generates a single product due to the steric effects of the residues in the binding pocket. Activation barriers were calculated by our recently developed environment-perturbed transition state sampling (EPTSS) method. Hydrophobic binding (the complementarity between substrate and enzyme), rather than electrostatic stabilization, is found to control the catalysis and selectivity of the PyrI4-catalyzed Diels–Alder reaction.

4.2 Introduction

The Diels–Alder (DA) reaction1 is one of the most versatile synthetic transformations for the construction of functionalized cyclohexenes. Various aspects of DA reactions, such as mechanism, selectivity, dynamics, catalysis, and applications in total synthesis have been extensively investigated since its original discovery ninety years ago.[2] Beginning in the late 1990s, natural enzymes that catalyze this reaction (Diels–Alderase or DAase) have been proposed and eagerly sought, due to the frequent occurrence of apparent DA adducts in secondary metabolites.[3] Generally, enzyme-catalyzed DA reactions are fast and stereoselective, and form only one of the four possible stereoisomeric products involving attack on either face of diene and dienophile. In

contrast, the uncatalyzed reactions are often slow and unselective. This difference has been instructive in the discovery of DAases.[4]

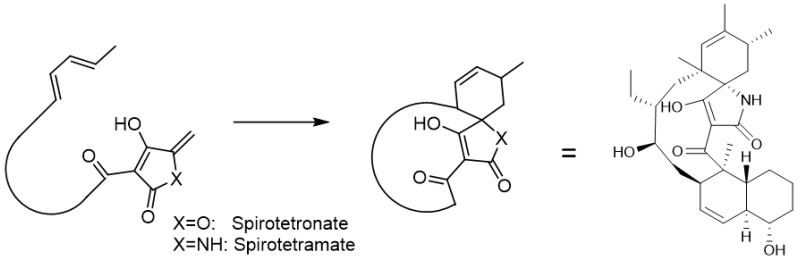
Various DAases have been identified and characterized in recent years, thanks to the technological advancement in whole-genome sequencing and searching tool of biosynthetic gene clusters.[5–9] Structural details about the first well-characterized monofunctional DAase, SpnF, were reported in 2015.10 Extensive computational studies were later reported to show the interplay between hydrogen bonding and hydrophobic cavity in promoting the reactivity and controlling the reaction dynamics.[11–15] Byrne et al. reported AbyU-catalyzed DA reaction and associated computational mechanistic studies. The DA reaction was proposed to involve a concerted, asynchronous transition state (TS).[16] However, the calculated barrier by the SCC-DFTB/ff14SB method largely overestimates the experimentally observed values, and no information on the origins of stereoselectivity was obtained. Tang and Houk reported the MycB-catalyzed DA reaction discovered in the biosynthesis of the myceliothermophins. We proposed the likely involvement of hydrogen bonding catalysis, which is important for the observed stereoselectivity.[9] This was supported computationally using a theozyme (QM cluster) model.

Enzymes can bind reactants and transition states by both electrostatic and hydrophobic interactions. Electrostatic preorganization has been widely invoked as the predominant factor in enzymatic catalysis for reactions involving charge separations, like methyltransferases, hydrolase, lyases, etc.[17] DAases, however, possess a hydrophobic pocket that allows the substrates diene and dienophile to fit in a reactive conformation to generate the DA adduct.[6] Charge separation is expected to be moderate in the process. This leads to an intriguing question regarding whether hydrophobic binding can override electrostatic interactions such as hydrogen bonding in

promoting the reactivity. Here, we would like to examine this hypothesis by studying the mechanism of a DA reaction catalyzed by PyrI4.

Recently, the groups of Pan and Liu reported the crystal structure of DAase PyrI4 in the biosynthesis of pyrroindomycins.[18] The pyrroindomycins belong to spirotetramate and spirotetronate class of natural products, and are promising antibiotics against drug-resistant bacterial pathogens. Through a combined experimental mutagenesis study and computational docking simulations, the authors presented a detailed picture of catalysis by PyrI4, including identification of loop motion that captures the substrate in a hydrophobic cavity and arranges the flexible side chains of spirotetramates into the geometry required to favor the formation of a single exo-DA adduct. They also drew attention to the acidic and hydrogen bond donor region of the site, which the tetramate occupies upon binding. Despite these important findings, no investigation of the transition states that are essential to catalysis and stereoselectivity was available.

Scheme 4.1. Intramolecular Diels-Alder reactions leading to spirotetronate or spirotetramate.



We have recently developed an efficient computational method, environment-perturbed transition-state sampling (EPTSS), to study single-molecule trajectories and to calculate activation barriers in enzymes. EPTSS has been used to elucidate the time-resolved mechanism of the SpnF-catalyzed DA reaction.[11] We now apply this method to investigate the PyrI4-catalyzed DA reaction in the biosynthesis of pyrroindomycins. The mode of catalysis and the origins of stereoselectivity were elucidated using combined quantum mechanics (QM) and molecular dynamics (MD) tools. Specifically, we have optimized the transition states of the four stereoisomeric Diels–Alder

reactions catalyzed by PyrI4, computed the energetics of these in the gas phase and in the enzyme, and evaluated the relative importance of hydrophobic shape-selective binding and hydrogen bonding on the catalysis by PyrI4. We find that the acceleration and selectivity caused by PyrI4 can be mainly attributed to the hydrophobic binding rather than hydrogen bonding.

4.3 Computational Methods

DFT Calculation:

Density functional theory computations were performed with Gaussian09. Geometry optimizations and frequency calculations were performed at the M06-2X/6-31G(d) level of theory, which has been shown to yield accurate energetics for cycloaddition reactions. Normal vibrational mode analysis confirmed that optimized structures were minima or transition states. Transition structures were connected to starting materials and products via IRC calculations to confirm the reaction pathway. Zero-point vibrational energy (ZPE) and thermal corrections were calculated using the unscaled M06-2X/6-31G(d) frequencies. Truhlar's quasiharmonic correction was used to compute molecular entropies to reduce error caused by the breakdown of the harmonic oscillator approximation, by setting all positive frequencies that are less than 100 cm–1 to 100 cm–1. M06-2X/6-311+G(d,p) single-point energies were computed on the optimized structures. Reported energies are Gibbs free energies. All 3D rendering of stationary points were generated using CYLview. GaussView and Avogadro were used to construct the structures used in our computations.

Conformational searches of the ground-state structures for all reactants, intermediates, and products were performed using the OPLS_2005 force field. A low mode/Monte Carlo (LMMC) search protocol optimized for sampling of macrocycles available in Macromodel 9.9 was

performed. Up to 150 lowest-energy structures in each case were then used as starting points for DFT optimization using M06-2X/6-31G(d).

Docking and Molecular dynamics

The conformations of the substrate and the transition structures optimized with the DFT method were docked into the enzyme active site using AutoDock Vina. Classical molecular dynamics (MD) were performed using the GPU code (pmemd) of the Amber 12 package on the substrate conformations and the transition states for 500 ns in enzyme. The FF99SBildn force field was used for the protein residues. Parameters for the substrate conformations and transition states were generated within the antechamber module using the general Amber force field (gaff), with the partial charges set to fit the electrostatic potential calculated at the HF/6-31G(d) level by the RESP model. The charges were computed according to the Merz-Singh-Kollman scheme using the Gaussian 09 package. Each enzyme complex was immersed in a pre-equilibrated truncated cuboid box with a 10 Å buffer of TIP3P water molecules using the tleap module. The systems were neutralized by addition of explicit counter ions (Na+ or Cl-). The systems were optimized for total 10000 steps, followed by gentle heating from 0 K to 300 K under constant-volume and periodic-boundary conditions. Each system was then equilibrated for 2 ns with a 2 fs time step. Production trajectories were then run for additional 500 ns under the same simulation conditions. During the classical MD on the transition states, restraining potentials of 500 kcal/mol/Å2 were applied to the reaction coordinates in the transition states.

Environment-perturbed Transition State Sampling (EPTSS)

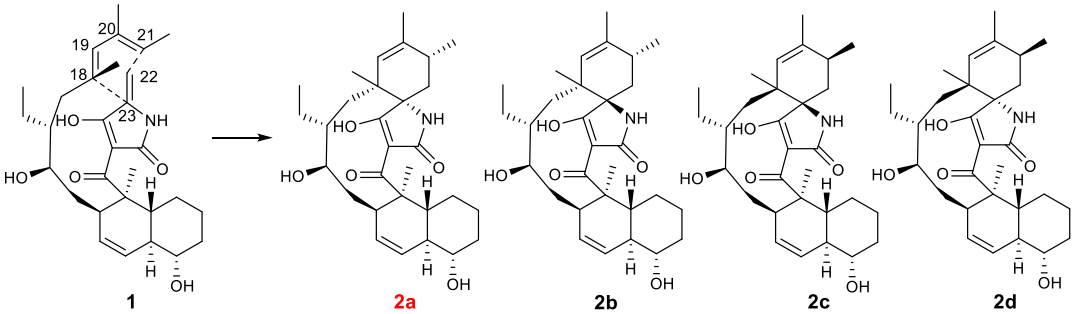
The Environment-perturbed Transition State Sampling (EPTSS) method proposed previously by our group was used to compute free energies of reactions in the enzyme. The method consists of three major steps. (1) Construction of the reaction/medium models. Initial structures of

transition structures TS-2a and TS-2b were optimized in the gas phase at the M06-2X/6-31G(d) level of theory, using Gaussian 09. For the reaction in enzyme, the substrate or the TS were docked into the enzyme active site using AutoDock Vina. Docking is detailed above. (2) Conformation sampling. Classical molecular dynamics (MD) was performed using Amber 14 on the complexes between PyrI4 and TS-2a or TS-2b. Fifty snapshots of TS-2a and TS-2b were sampled from production MD runs at 1 ns intervals in PyrI4. (3) Free energy calculations. For each snapshot, the coordinates of the environment (PyrI4) were frozen, the geometry of TS-2a and TS-2b was optimized with QM/MM with M06-2X/6-31G(d) as QM, and the free energy was computed from harmonic vibrational partition functions at the stationary point. Free energy difference between TS-2a and TS-2b were computed as ⟨E(TS-2b)⟩–⟨E(TS-2a)⟩, where angle brackets indicate an ensemble average over 50 snapshots.

4.4 Results and Discussion

The Diels–Alder reaction of substrate 1 can produce four possible products, 2a to 2d, depending on which face of the diene and alkene is attacked Products 2a and 2c are exo and products 2b and 2d are endo, where endo refers to the position of the enol double bond relative to diene carbons 19 and 20 in the TS. Pyrl4 converts compound 1 exclusively to product 2a. We carried out DFT calculations to compare the energy barriers leading to the four different products and to determine the intrinsic selectivity in the Diels–Alder reaction of 1.

Scheme 4.2. Four possible selective products in the Diels–Alder reaction of 1.



The energetics of the reaction of substrate 1 leading to product 2a in the gas phase are shown in Figure 1. In the global minimum (conf-GM), the 1,3-diene and alkene groups are over 10 Å apart (SI). Substrate 1 needs to undergo a large conformational change to be organized for the Diels–Alder reaction. Four reactive conformations leading to the four stereoisomeric products were also located and optimized (SI). Reactive conformations 2a-d (RC 2a-d) are the conformers leading to products 2a-d respectively. The reactive conformations are 6.5─10.1 kcal/mol higher in energy than conf-GM.

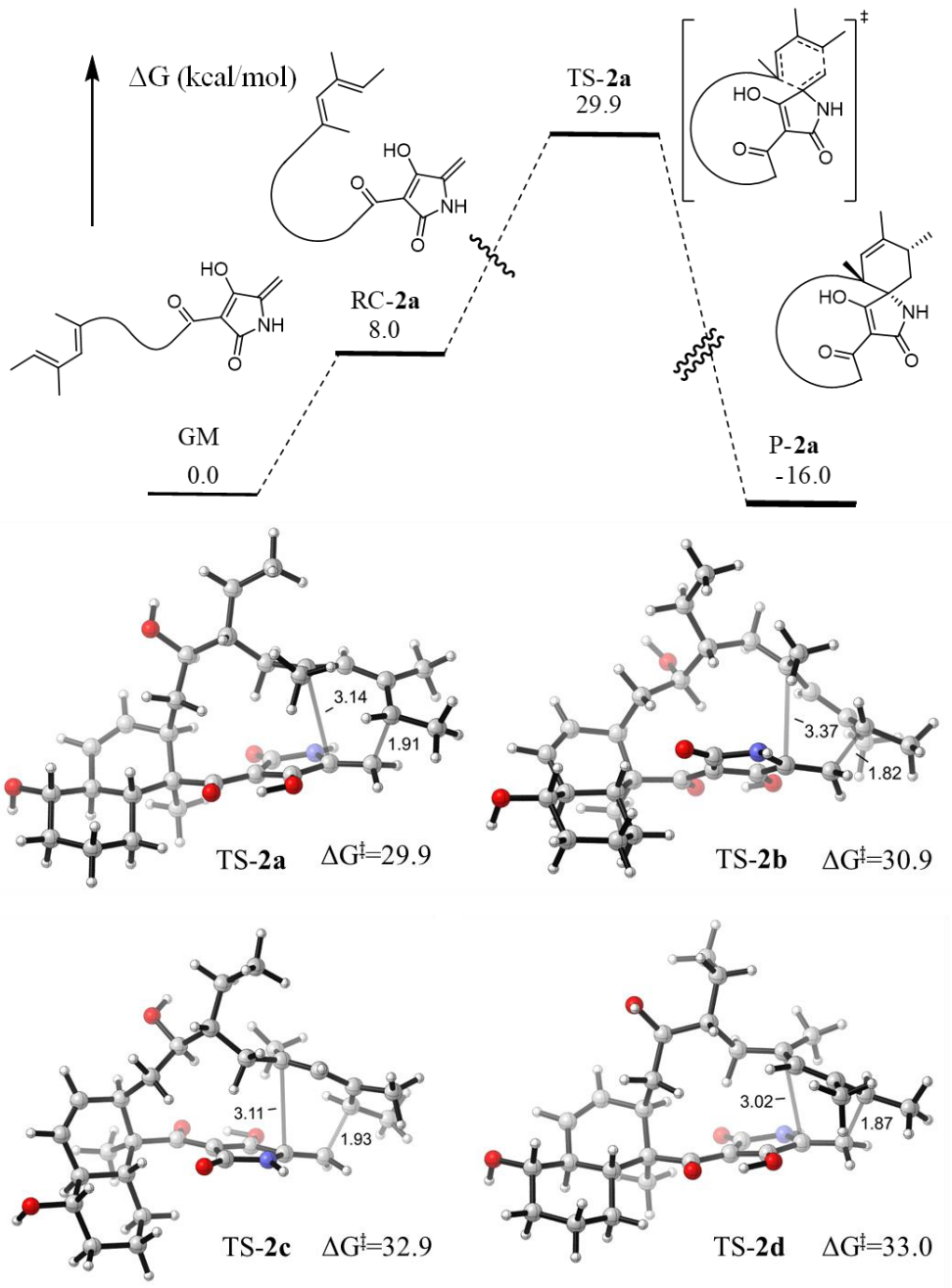


Figure 4.1. The energetics of the Diels–Alder reaction and the transition states of the four stereoisomeric products in the gas phase.

The four transition states TS 2a-d leading to products P 2a-d are shown in Figure 1. All transition states are highly asynchronous transition states with the terminal forming bond length of 1.8-2.0 Å, and the second forming bond length of 3.0-3.4 Å. The high barriers indicate that the uncatalyzed reaction is very slow at room temperature. This is partially due to the fact that the

substrate needs to undergo large conformational changes to achieve the reactive conformations. Once the reactive conformation is achieved the additional barriers are lower (21.9-26.1 kcal/mol) because of the complementary electron-rich diene and electron-deficient dienophile. TS-2a has the lowest energy, while TS-2b is only 1 kcal/mol higher in energy. The gas phase barriers predict that the intrinsic selectvities at room temperature would be 83.6 : 15.5 : 0.5 : 0.4 for 2a–d, respectively.

We explored how PyrI4 operates first by docking conf-GM and four RC 2a-d of substrate 1, as well as the four transition structures (TS 2a-d) into the crystal structure of the PyrI4 enzyme with Autodock Vina. The binding affinity predicted by Autodock Vina indicates that RC 2a-2d has higher binding affinity than conf-GM (Table 1). TSs general has better affinity than their responding RCs, and among them TS-2a binds the best to PyrI4. Figure 2 shows the binding of conf-GM, 2a and 2b in the PyrI4 binding pocket. In the front view, the tetramate group binds in the right side while the diene group binds in the left side in conf-GM, 2a and 2b. The diene group in conf-GM is extending outside of the binding pocket next to the α0 lid, which is found experimentally to be essential to substrate binding and catalysis. The RC and TS for 2a are bound in nearly identical ways. The orientation of the tetramate dienophile is reversed relative to residue Q115 in exo (2a) and endo (2b). Q115 is next to the hydroxyl and carbonyl functionalities in the tetramate of conf-GM, 2a and 2d (Docking results in SI). In 2b and 2c, Q115 is near the amine and carbonyl groups in the tetramate, forming hydrogen bonds with both.

Table 4.1. Binding affinities (kcal/mol) of RCs, TSs, and Conf-GM in PyrI4.

Binding Affinity	2a	2b	2c	2d	Conf-GM
RC	-10.4	-8.8	-9	-9.4	-7.8
TS	-12.6	-10.2	-9.3	-9.7	-

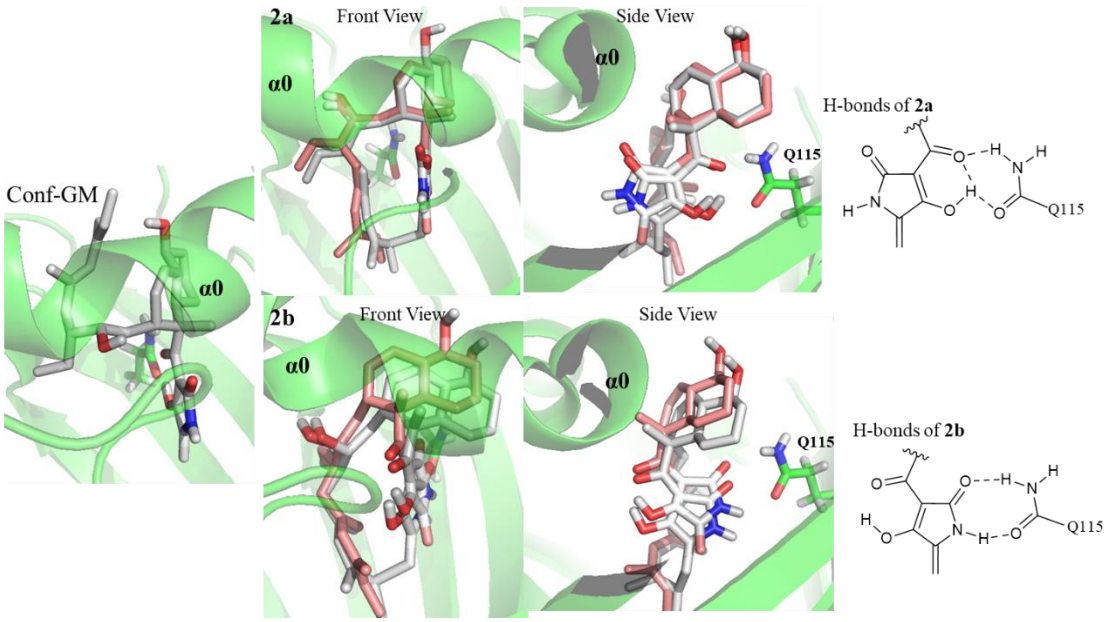


Figure 4.2. Reactive conformations and transition states 2a and 2b docked into the crystal structure of PyrI4. Reactive conformations are shown in red; transition structures in grey.

To evaluate if the hydrogen bonds accelerate the Diels–Alder reaction, we computed the energy barriers of the four TSs in a theozyme model using formamide as a model for Q115 (Figure 3). Exo-TSs, TS-2a and TS-2d form hydrogen bonds with formamide with the carbonyl and hydroxyl groups in the tetramate, while endo TSs, TS-2b and TS-2c, each form hydrogen bonds with the amide moiety of the tetramate instead. Hydrogen bonding to the amide group is clearly favored, and the formation of the endo adduct 2b would be accelerated most (by 2-5 kcal/mol) if both hydrogen bonds could occur in the enzyme. The energy of TS-2a/2d is lowered by only 1-2 kcal/mol by hydrogen bonding with formamide. The stabilization of TS-2a/2d is less significant because the hydrogen bonds between the carbonyl and hydroxyl group of the tetramate and the formamide is weaker compared to those in TS-2b/2c.

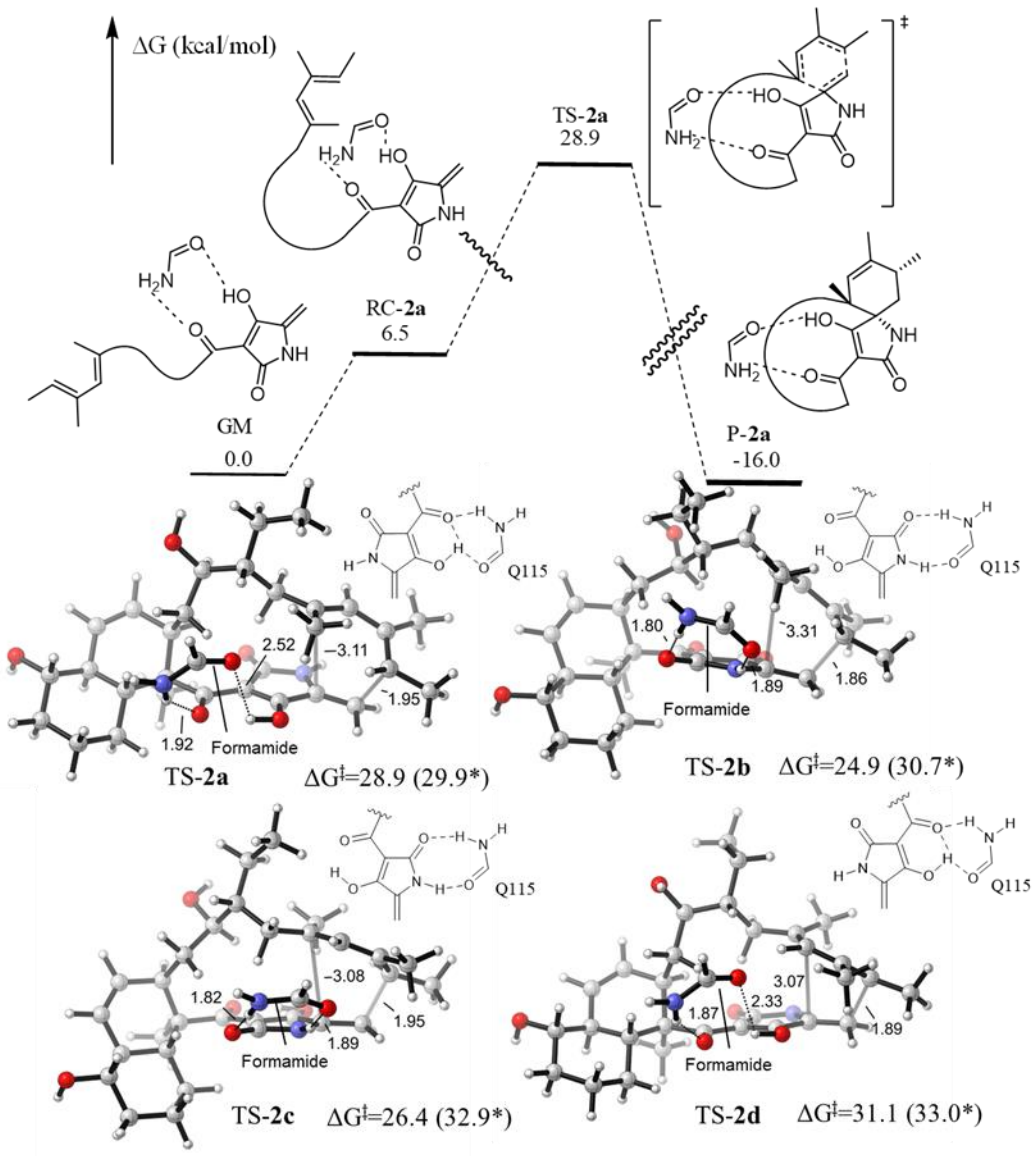


Figure 4.3. R Energetics of the Diels–Alder reaction of 1 in a theozyme model for hydrogen bonds. A formamide replaces the glutamine in the enzyme. Energy barriers of each TS in the gas phase are labeled with *.

To evaluate the influence of the hydrophobic pocket on the DA reaction, we further carried out 500 ns MD simulations starting from the four RCs and four TSs docked into the enzyme. Classical MD simulations of substrate 1 (starting from RC 2a-2d, TS 2a-2d) in Pyrl4 and in water (starting from RC-2a) were carried out. In water, the conformation of the substrate is mostly the global minimum conformation (conf-GM). The distances from diene (C21) to dienophile (C22) are plotted during the 500 ns MD simulations in Figure 4. The distance from C21 to C22 is over

10 Å on average. In Pyrl4, RC-2a is maintained as a reactive conformation for the DA reaction, as the 1,3-diene and the dienophile groups remain close to each other. The distance from C21 to C22 in 2a is average 4 Å for 500 ns. Starting from the other reactive conformations (RC 2b-d) in the enzyme, the distance between C21 and C22 is shifted to 6-12 Å. We compare the representative snapshots from MD simulation of RCs. In RC-2a, the dienophile forms hydrogen bonds with Q115 and the diene group is kept in proximity to the dienophile by surrounding hydrophobic residues. In RC-2b, the hydrogen bonds between Q115 and tetramate dienophile are not formed. In RC-2b/2c, the diene group binds in a different orientation than RC-2a, which places C21 closer to the Y177. In RC-2d, the diene group unbinds from the hydrophobic pocket. As a result, the diene and dienophile groups are apart in 2b/2c/2d, and they deviate from the original reactive conformations. We also carried out 500 ns MD simulation with TSs bound in PyrI4. All TSs (2a-2d) are constrained with harmonic constraints to keep the TS geometry. Representative snapshots from MD simulation are shown in Figure 4. We compute the change in solvent accessible surface area (SASA) for TS-PyrI4 complex. TS-2a has the largest decrease in SASA as it binds into PyrI4, which indicates that TS-2a benefit the most from hydrophobic effect. We conclude that the shape of the binding pocket of PyrI4 is complementary to 2a (both RC and TS), and stabilizes it through hydrophobic substrate binding.

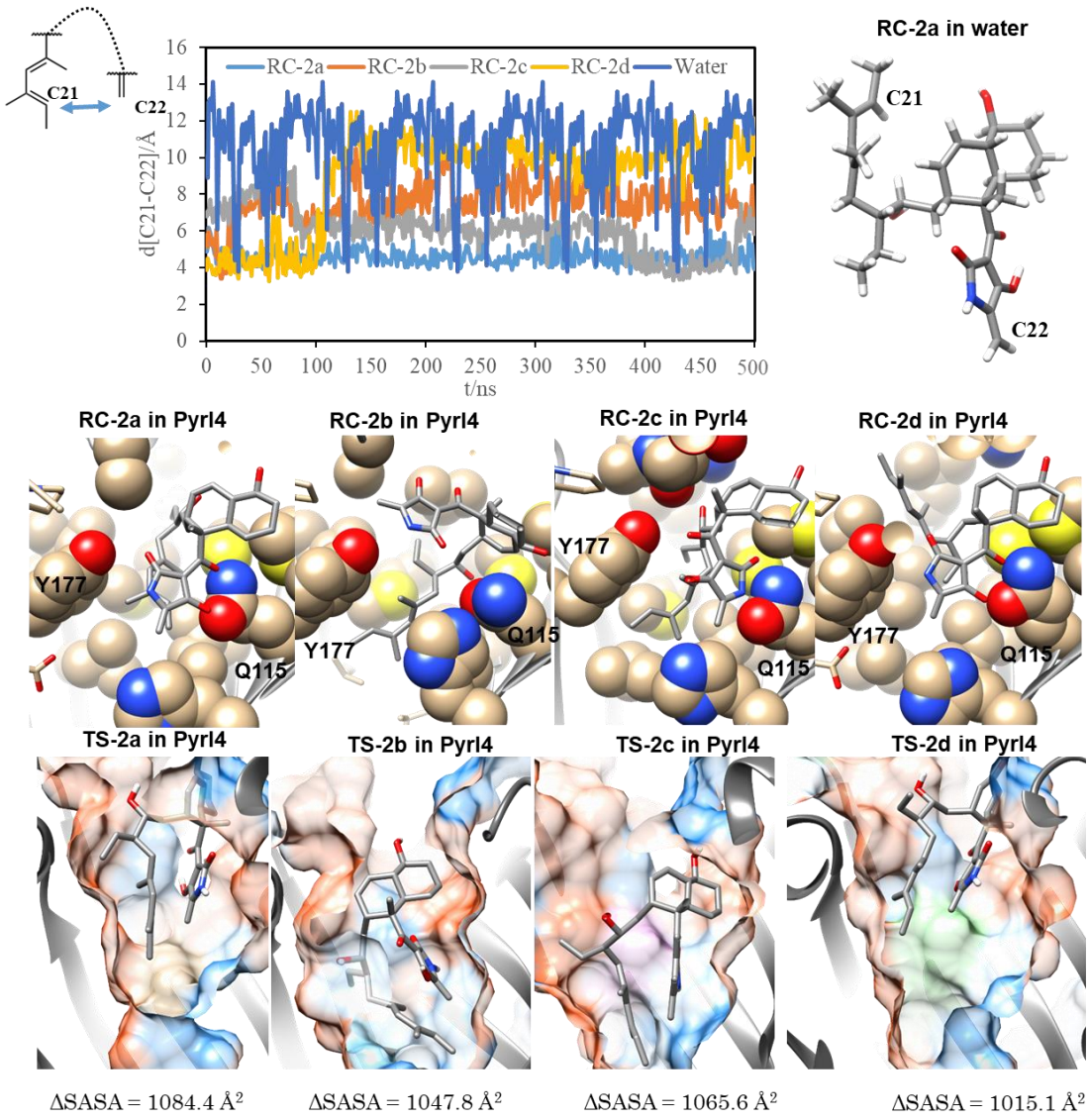


Figure 4.4. d[C21-C22] for MD simulations of PyrI4 with different starting conformations.

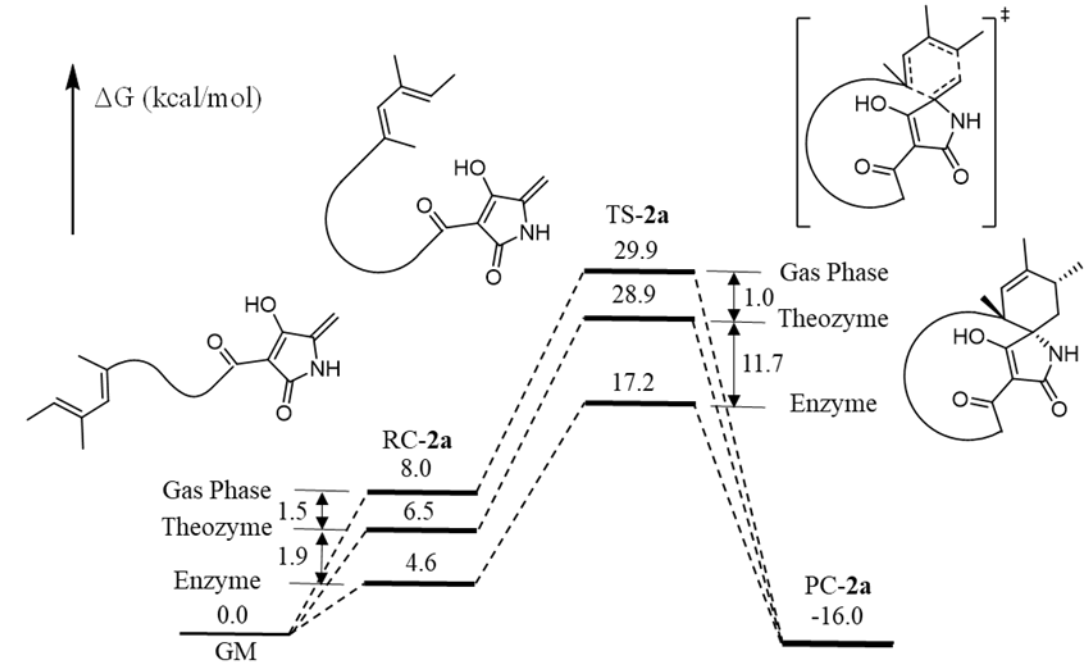
In the theozyme model, we have shown that hydrogen bonds with formamide favor TS-2b by 4.0 kcal/mol compared to TS-2a. However, the MD simulation indicates that RC-2a is most favorable in terms of substrate binding and pre-organization. To further investigate the stereoselectivity in the enzyme, EPTSS was applied to compute the energy barriers of TS-2a and TS-2b in PyrI4. The computed energy barrier for TS-2b in PyrI4 is 21.8 kcal/mol, which is 3.1 kcal/mol lower than theozyme computation (Table 1). The energy barrier of TS-2a in the enzyme is 17.2 kcal/mol, which is very close to the experimental energy barrier of 16.7 kcal/mol computed

by the Eyring equation from the experimental rate constant. The energy of TS-2a is much lower than TS-2b in PyrI4. The large energy difference between TS-2a and TS-2b is also consistent with the exclusive selectivity for product 2a in experiment with PyrI4.

Table 4.1. Energy Barriers Computed with Different Models and Comparison to Experiment.

	Gas Phase	Theozyme	Enzyme	Experiment[a]
TS-**2a**	**29.9 (20.8[b])**	28.9 (21.7)	**17.2±2.8**	16.7
TS-**2b**	30.9 (21.5)	**24.9 (18.5)**	21.8±1.7	-

a. Calculated with Eyring TS theory from the experimental rate of 342.6 min^{-1}
b. Energy in SMD implicit water model



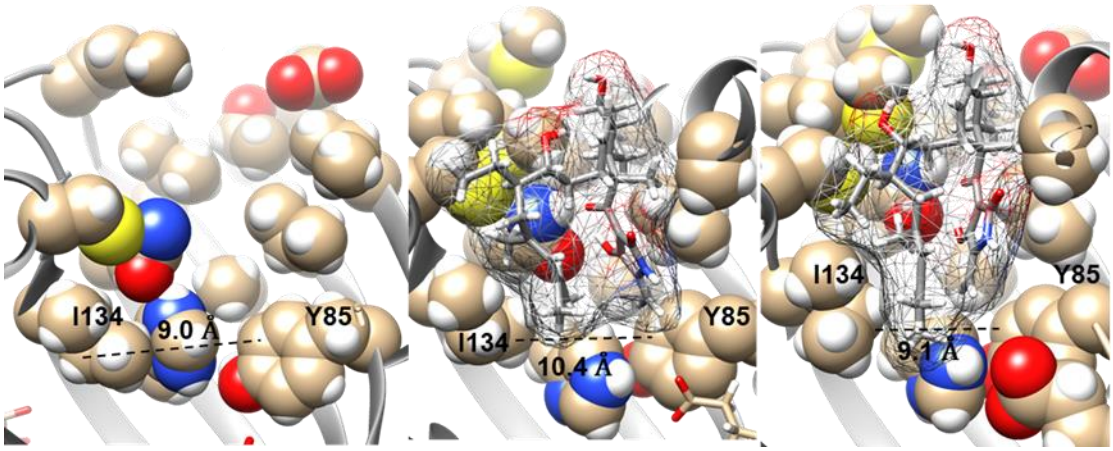


Figure 4.5. Comparison of reaction pathway in the gas phase, theozyme and enzyme

Calculations reveal the mechanism of TS stabilization by PyrI4, and the contribution from substrate binding and hydrogen bonds (Figure 5). The theozyme model proposes that the energy of TS-2a is lowered by 1 kcal/mol compared to the gas phase, which is attributed to hydrogen bonding. Experimentally, Wen Liu et al. found that mutation Q115A, which eliminates the hydrogen bonding with substrate, leads to loss of only 2/3 of the PyrI4 activity18. The partial loss of activity is attributable to the minor increment of energy barrier of the reaction by 1 kcal/mol in the absence of hydrogen bonding. In the enzyme, the energy of TS-2a is about 10 kcal/mol less than in the theozyme. The large energy change in the enzyme is mainly through substrate binding and hydrophobic interactions with the residues in the binding pocket, as there are no electrostatic interactions other than the hydrogen bonds modeled in the theozyme (SI).

PyrI4 also stabilizes the reactive conformation RC-2a by 3.4 kcal/mol, which is significantly smaller than the stabilization of TS. We compared the binding of RC-2a and TS-2a in PyrI4 binding pocket with the apo structure of PyrI4. The diene and dienophile are in hydrophobic contact with I134 and Y85, which form the narrow bottom of the binding pocket, with the distance (I134 CG1—Y85 CE2) to be 9.0 Å in PyrI4 apo structure. When RC-2a binds in PyrI4, the average distance between the two residues in MD simulation is 10.4 Å. In TS-2a and PyrI4 complex, the average distance is 9.1 Å. TS-2a fits in the binding pocket better because diene and dienophile are much closer to each other. Therefore, the catalysis and stereoselectivity of PyrI4 originates mainly from TS stabilization, and that is by hydrophobic binding of an active site that is highly complementary to the TS-2a geometry.

4.5 Conclusion

In summary, we explored the mechanisms and selectivities of the PyrI4 catalyzed DA reaction. The DFT calculations in the gas phase reveal that the substrate must achieve a TS-like conformation that is 6.5-10 kcal/mol above the global minimum. The reaction has a slight intrinsic preference for the experimentally observed exo product in the gas phase. MD simulations of the enzyme provide insights into the mechanism of catalysis by PyrI4. Hydrogen bonds with Q115 stabilize all 4 TSs, but the preference for 2a occurs through more favorable van der Waals interactions with TS-2a. These insights provide a roadmap to engineer Diels–Alderases to make stereo-isomeric products not observed in Nature.

4.6 References

(1) Diels, O.; Alder, K. Synthesen in Der Hydroaromatischen Reihe. Justus Liebigs Ann. Chem. 1928, 460 (1), 98–122.

(2) Nicolaou, K. C.; Snyder, S. A.; Montagnon, T.; Vassilikogiannakis, G. The Diels–Alder Reaction in Total Synthesis. Angew. Chem. Int. Ed. 2002, 41 (10), 1668–1698.

(3) Stocking, E. M.; Williams, R. M. Chemistry and Biology of Biosynthetic Diels–Alder Reactions. Angew. Chem. Int. Ed. 2003, 42 (27), 3078–3115.

(4) Sato, M.; Yagishita, F.; Mino, T.; Uchiyama, N.; Patel, A.; Chooi, Y.-H.; Goda, Y.; Xu, W.; Noguchi, H.; Yamamoto, T.; et al. Involvement of Lipocalin-like CghA in Decalin-Forming Stereoselective Intramolecular [4+2] Cycloaddition. ChemBioChem 2015, 16 (16), 2294–2298.

(5) Minami, A.; Oikawa, H. Recent Advances of Diels–Alderases Involved in Natural Product Biosynthesis. J. Antibiot. (Tokyo) 2016, 69 (7), 500–506.

(6) Jeon, B.; Wang, S.-A.; Ruszczycky, M. W.; Liu, H. Natural [4 + 2]-Cyclases. Chem. Rev. 2017, 117 (8), 5367–5388.

(7) Klas, K.; Tsukamoto, S.; Sherman, D. H.; Williams, R. M. Natural Diels–Alderases: Elusive and Irresistable. J. Org. Chem. 2015, 80 (23), 11672–11685.

(8) Zheng, Q.; Tian, Z.; Liu, W. Recent Advances in Understanding the Enzymatic Reactions of [4+2] Cycloaddition and Spiroketalization. Curr. Opin. Chem. Biol. 2016, 31, 95–102.

(9) Li, L.; Yu, P.; Tang, M.-C.; Zou, Y.; Gao, S.-S.; Hung, Y.-S.; Zhao, M.; Watanabe, K.; Houk, K. N.; Tang, Y. Biochemical Characterization of a Eukaryotic Decalin-Forming Diels–Alderase. J. Am. Chem. Soc. 2016, 138 (49), 15837–15840.

(10) Fage, C. D.; Isiorho, E. A.; Liu, Y.; Wagner, D. T.; Liu, H.; Keatinge-Clay, A. T. The Structure of SpnF, a Standalone Enzyme That Catalyzes [4 + 2] Cycloaddition. Nat. Chem. Biol. 2015, 11 (4), 256–258.

(11) Yang, Z.; Yang, S.; Yu, P.; Li, Y.; Doubleday, C.; Park, J.; Patel, A.; Jeon, B.; Russell, W. K.; Liu, H.; et al. Influence of Water and Enzyme SpnF on the Dynamics and Energetics of the Ambimodal [6+4]/[4+2] Cycloaddition. Proc. Natl. Acad. Sci. 2018, 201719368.

(12) Zheng, Y.; Thiel, W. Computational Insights into an Enzyme-Catalyzed [4+2] Cycloaddition. J. Org. Chem. 2017, 82 (24), 13563–13571.

(13) Jeon, B.; Ruszczycky, M. W.; Russell, W. K.; Lin, G.-M.; Kim, N.; Choi, S.; Wang, S.-A.; Liu, Y.; Patrick, J. W.; Russell, D. H.; et al. Investigation of the Mechanism of the SpnF-Catalyzed [4+2]-Cycloaddition Reaction in the Biosynthesis of Spinosyn A. Proc. Natl. Acad. Sci. 2017, 114 (39), 10408–10413.

(14) Medvedev, M. G.; Zeifman, A. A.; Novikov, F. N.; Bushmarinov, I. S.; Stroganov, O. V.; Titov, I. Y.; Chilov, G. G.; Svitanko, I. V. Quantifying Possible Routes for SpnF-Catalyzed Formal Diels–Alder Cycloaddition. J. Am. Chem. Soc. 2017, 139 (11), 3942–3945.

(15) Andes Hess, B.; Smentek, L. Concerted, Highly Asynchronous, Enzyme-Catalyzed [4 + 2] Cycloaddition in the Biosynthesis of Spinosyn A ; Computational Evidence. Org. Biomol. Chem. 2012, 10 (37), 7503–7509.

(16) Byrne, M. J.; Lees, N. R.; Han, L.-C.; van der Kamp, M. W.; Mulholland, A. J.; Stach, J. E. M.; Willis, C. L.; Race, P. R. The Catalytic Mechanism of a Natural Diels–Alderase Revealed in Molecular Detail. J. Am. Chem. Soc. 2016, 138 (19), 6095–6098.

(17) Shurki, A.; Štrajbl, M.; Villà, J.; Warshel, A. How Much Do Enzymes Really Gain by Restraining Their Reacting Fragments? J. Am. Chem. Soc. 2002, 124 (15), 4097–4107.

(18) Zheng, Q.; Guo, Y.; Yang, L.; Zhao, Z.; Wu, Z.; Zhang, H.; Liu, J.; Cheng, X.; Wu, J.; Yang, H.; et al. Enzyme-Dependent [4 + 2] Cycloaddition Depends on Lid-like Interaction of the N-Terminal Sequence with the Catalytic Core in PyrI4. Cell Chem. Biol. 2016, 23 (3), 352–360.

Chapter 5. Studying the Substrate Specificity and Reactivity of MycG: A Multifunctional Enzyme

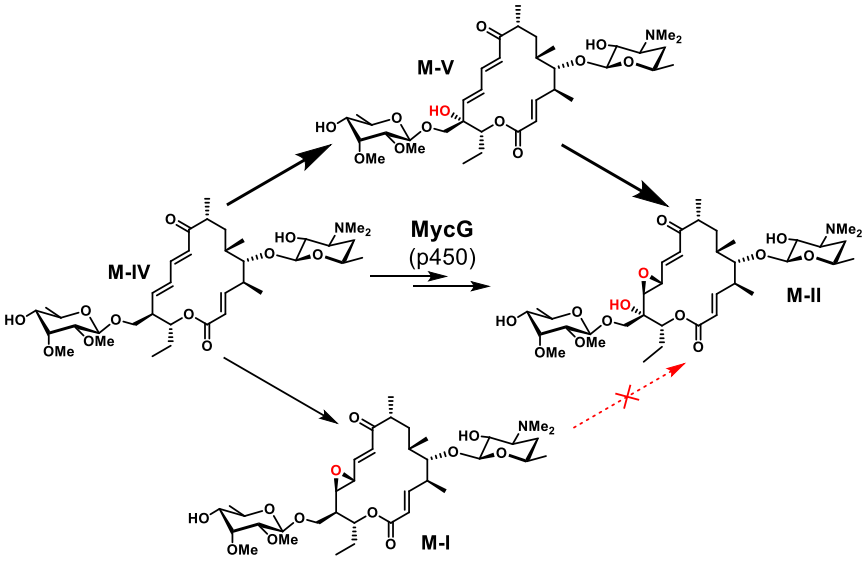
5.1 Abstract

Molecular dynamics are well known to reveal the dynamics information beyond static X-ray structures. In this study, we show that molecular dynamics of bifunctional P450 MycG explains the selectivities to the natural substrate M-IV, and predict mutations that improve reactivities to the unreactive substrates (M-VI and M-III with 99% identity in structure with M-IV). Conformation search of substrate in enzyme using restrained MD clarify the interactions during hydroxylation and epoxidation. The deoxysugar of all substrates reside in a hydrophobic cavity composed of three leucines. The polarity of the deoxysugars effects their affinity to the hydrophobic cavity and also the gating between R75 and E173 in the FG and BC loops, which further enlarges the difference between three substrates. MD simulations on possible mutants predict that enhancing the gating between R75 and E173 is beneficial to improved reactivity for M-VI and M-III. Experimental data further confirms that the yield of oxidation products for M-III and M-VI is doubled or quadrupled respectively. These findings demonstrate the power of molecular dynamics in predicting enzymatic reactions even when the difference between substrate.

5.2 Introduction

P450 enzymes catalyze a wide array of reactions including, but not limited to, hydroxylation, epoxidation, C-C coupling, N-dealkylation, sulfoxidation, and arene oxidation.[1–6] While enzymes have evolved to catalyze reactions with excellent selectivity and efficiency, a few are able to catalyze multiple oxidative steps involving separate reaction mechanisms. MycG is one such enzyme which catalyzes the sequential hydroxylation and epoxidation of M-IV to M-II.[7–9]

The order of these oxidative steps is critical. M-I is a side product in which the epoxide is installed first. If M-I is fed back to MycG, no hydroxylation to M-II is observed.



Scheme 5.1. MycG catalyzed oxidation of M-IV. Hydroxylation to M-V followed by epoxidation leads to M-II. Epoxidation of M-IV leads to shunt product M-I.

In addition to selectively performing 2 sequential and distinct oxidative steps, MycG is highly substrate specific. M-IV, M-III, and M-VI are identical in macrocyclic structure and differ only in extent of methylation of the mycinose sugar moiety. Despite having ≥110 atoms in common, >94% identical to M-IV, M-III and M-VI experience very low levels of oxidation when tested with MycG. Based on LC-MS and MS-MS data, it is believed that these minor products are hydroxylation of the respective macrocycles but the exact structure of the oxidized products could not be determined since there was such a small amount of product formed.

We set out on a computational investigation of the profound substrate selectivity and reactivity preferences of MycG. Molecular dynamics (MD) simulations and Metadynamics were used to determine how these minor differences in the sugar at C14, remote from the sites of oxidation, have an impact on activity. To determine why the order of oxidation of M-IV is so

critical, we used density functional theory (DFT) calculations. Collectively, these methods complement each other to fully understand this complex, multi-functional enzyme.

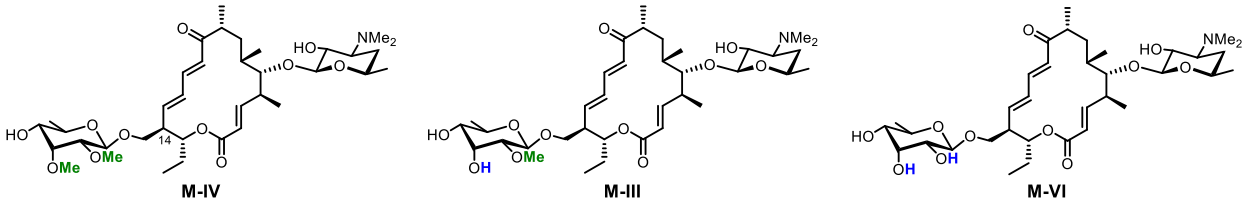


Figure 5.1. Macrocycles in the biosynthesis of M-II which differ in the extent of methylation of deoxysugar at C14. Only M-IV is effectively oxidized by MycG.

5.3 Computational Methods

QM.

DFT calculations were performed using the Gaussian09 suite.[10] Conformational sampling of substrates and transition states were performed using Spartan08 and Spartan14. The default settings for torsional sampling were used. For transition states, the O-H-C bonds and were frozen and other substrate torsions were sampled, including rotation about the C-H bond being abstracted. The iron-oxo model and transition states were computed at the quartet spin state. Geometry optimizations and frequency calculations were performed at the B3LYP level. The 6-31G(d) basis set was used on all atoms except Fe, for which the LANL2DZ pseudopotential was used. Saddle point transition structures were confirmed by the presence of one imaginary frequency corresponding to the desired transformation. The quasi-harmonic correction, as described by Truhlar, was used to adjust the Gibbs free energy. Single point calculations were performed at the B3LYP-D3(BJ) level with CPCM implicit solvation with water. The LANL2DZ pseudopotential was used for Fe and 6-311+G(d,p) was used for all other atoms.

MD simulations.

The heme iron(IV)-oxo complex involved in the cytochrome-catalysed oxidative hydroxylation cycle (compound I) was used to model the active form of the MycG cofactor. MD simulations were performed using the GPU code (pmemd) of the AMBER 14 package.[11–13] The Amber-compatible parameters developed by Cheatham et al. were used for compound I and its axial Cys ligand.[14] The parameters for the substrates (M-IV, M-III, M-VI) were generated with the antechamber module using the general AMBER force field (gaff),[15,16] with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model.[17] The charges were calculated according to the Merz-Singh-Kollman scheme using the Gaussian 09 package. MycG was immersed in a pre-equilibrated truncated octahexon with a 10 Å buffer of TIP3P water molecules using tleap module. All subsequent calculations were done using Amber/ff99SBildn force field. We did 100 ns restrained simulations starting from the crystal structure (co-crystallized with M-IV) of MycG (PDB 2Y98). M-III and M-VI were manually docked into MycG with the same binding pose as M-IV. 500 kcal/(mol·Å)2 harmonic restraints were added on the HC14--OFe distance. The system was optimized and then gradually heated to 300 K, and was then allowed to equilibrate. Production trajectories were run for 100 ns under restraint and then 400 ns without any restraint.

Metadynamics.

Metadynamics were carried out to quantify the difference in MycG activities to each substrate. NAMD v2.9 was used for the Metadynamics simulations with ff99SBild force fields.[18,19] Well-tempered Metadynamics in conjunction with multiple-walker algorithm were used to accelerate the convergence of the free energy landscape computed from the Metadynamics simulations. The Langevin thermostat controlled at 300 K and Langevin barostat kept the pressure

at 1 atm. We chose the H14--OFe distance as the collective variable to describe the binding of substrate in MycG, with the range of the collective variable to be 7.9 – 1.1 Å.

5.4 Results and Discussion

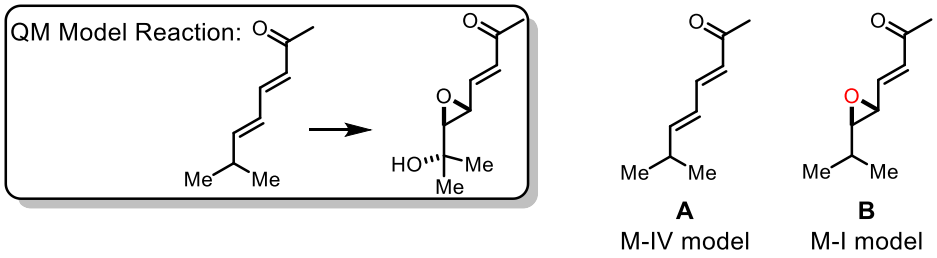


Figure 5.2. Model reaction used to study the activity of MycG with DFT.

Using A as a model for M-IV and B as a model for M-I, we performed density functional calculations to determine the barriers to C-H abstraction. With respect to separated reactants, the free energy barrier to C-H abstraction from A is 14.6 kcal/mol. The barrier to abstraction from B is 20.1 kcal/mol. This 5.5 kcal/mol increase in barrier for abstraction from B means that abstraction from the epoxidized substrate is 10,000 times slower than abstraction from A. This drastic difference in activity is in accord with the lack of activity toward M-I. The presence of the epoxide in M-I decreases the inherent reactivity, so that MycG is unable to hydroxylate this intermediate.

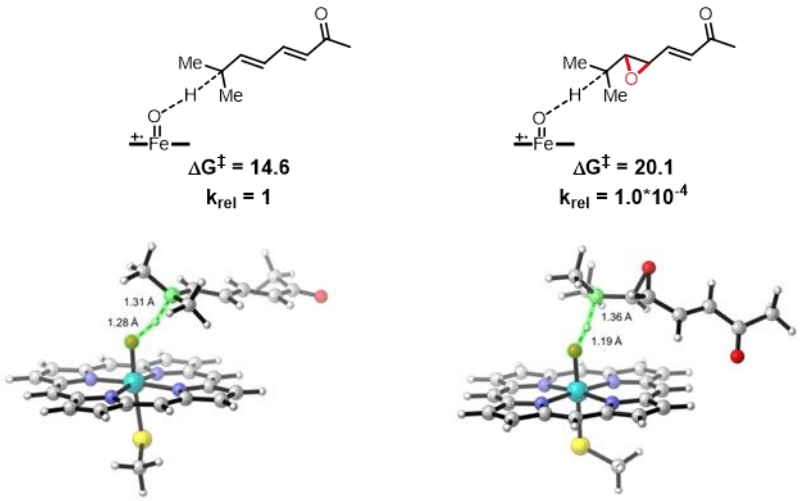


Figure 5.3. Free energies and transition structures of the C-H abstraction from A and B.

There are 2 attributes of the installed epoxide which could destabilize the developing radical and be the origin of the barrier increase: 1) the lack of allylic stabilization; 2) the electron-withdrawing nature of the oxygen. To probe these two properties of the epoxide independently we computed the barriers to C-H of two additional substrate models. Model C mimics the electron withdrawing nature of the oxygen in M-I but still has the conjugation present as in M-IV. Model D mimics the loss of conjugation due to the epoxide but no electron withdrawing groups are present. The barriers to C-H abstraction from C and D are 15.9 and 18.9 kcal/mol respectively. In comparison to the abstraction from A, the barrier increases by +1.3 kcal/mol for C and +4.3 kcal/mol for D. Both factors therefore contribute to the barrier increase in B but lack of allylic stabilization is the most significant factor.

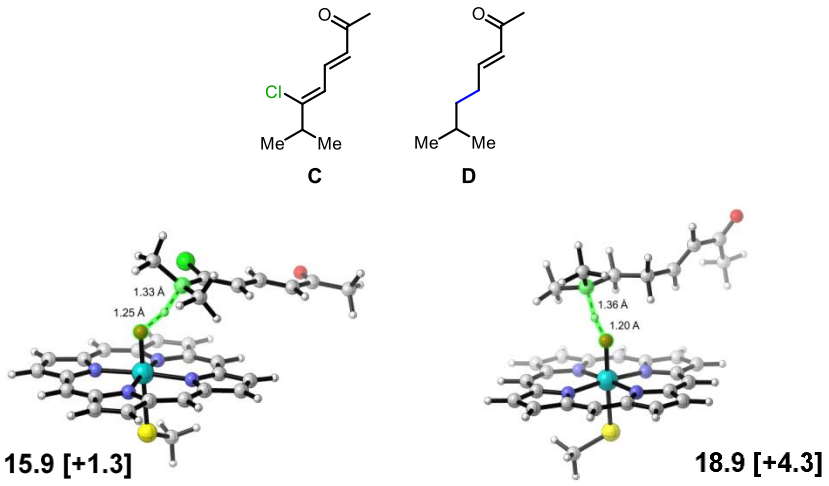


Figure 5.4. Top: substrate models C and D; Bottom: ΔG‡[difference from A] for C-H abstraction from C(left) and D(right).

Restrained MD simulations. The crystal structure of MycG with M-IV in the heme domain shows that M-IV binds in a "mycinose-in desosamine-out" binding mode. The mycinose sugar in M-IV is above the heme cofactor, hindering the macrolactone's access to the iron. The site of hydroxylation site of M-IV is about 8.9 Å away from the heme iron, which is too far to allow for activity. Anovel binding mode was revealed by 100ns restrained MD simulations with harmonic

restraints added along the distance from C14 to the oxygen of the iron-oxo compound. The restrained complex led to geometries more like that expected for H-abstraction (see Figs. 3 & 4). The restrained MD simulation illuminated a different binding mode characterized by a much shorter C14 to O=Fe distance of 3.5 Å (see Figure 5). The macrolactone ring's orientation is rotated by ~90 degrees, with the deoxysugar pushed into a hydrophobic cavity composed of L227, L83 and L94 residues. The desosamine sugar is also pulled toward the heme and, as a result, the two salt bridge interactions with E173 and E77 that are loosely shown by the crystal structure are strengthened.

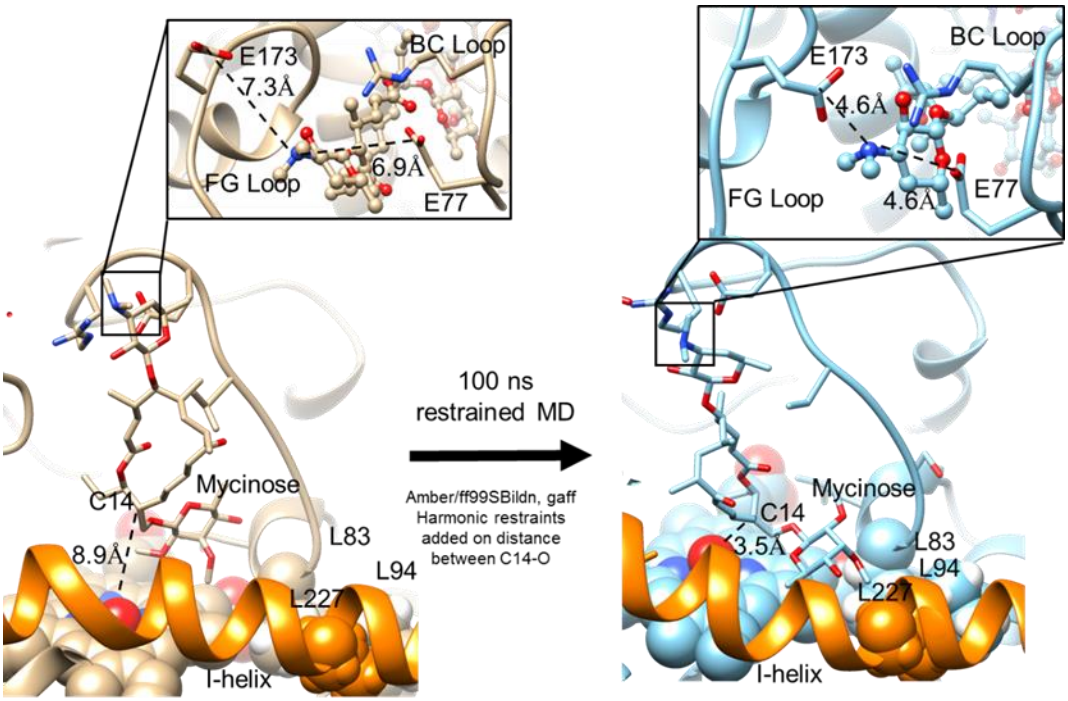


Figure 5.5. Binding configuration of M-IV as in crystal structure (left) and binding configuration with C14 constrained to be close to the heme iron (right). FG (with E173 residue on it) and BC loops (has E77 residue) are two loops upon the binding pocket that are close to each other. I-helix (where L227 locates) are the α-helix next to the heme group.

Metadynamics. To compare the free energies of the two different binding modes, Metadynamics were carried out on M-IV, M-III, and M-VI to calculate the free energies of the two binding modes. 250 ns metadynamics calculations were run using Amber/ff99SBildn force field.

The free energy profiles were plotted along the HC14--OFe distance from 7.9 Å to 1.1Å, as shown in Figure 6. The free energy profiles show there are two minima for each substrate within MycG. One minimum is at the HC14--OFe distance of ~7.9 Å, representing the non-reactive binding configuration (similar to the crystal structure). The other minimum is at the distance of ~2.7 Å, corresponding to a reactive binding configuration. An energy barrier must be overcome for C14 to get close to iron-oxo compound, but this barrier is only 5-7 kcal/mol depending on the substrate, in Figure 6. These two minima are essentially isoenergetic for M-IV. The reactive binding pose are less stable than the crystal structure-like pose for M-III and M-VI.

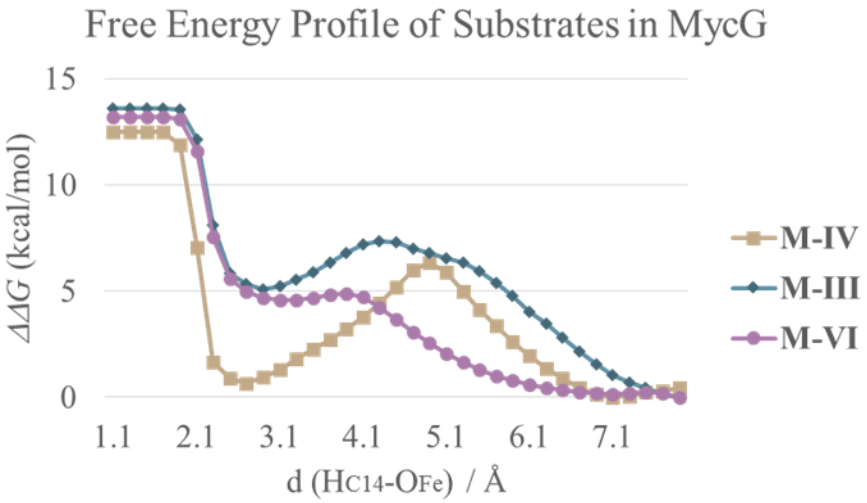


Figure 5.6. Free energy profile of M-IV (orange), M-III (blue), M-VI (purple) binding in MycG along the HC14--OFe bond.

Substrate recognition. With this information in hand, unrestricted MD simulations were performed on the active bind poses of M-IV, M-III, and M-VI. As mentioned previously, the only difference among these three substrates is the sugar attached to C14. In M-IV this sugar is mycinose, 6-deoxyallose in M-VI, and javose in M-III. As determined by the constrained MD, this sugar sits in a hydrophobic pocket of L227, L94 and L83 when in the active binding pose. The metadynamics show that in this active pose, the MycG+M-IV complex is just as stable as in the crystal structure pose. Alternatively, in this active pose, M-III and M-VI destabilized from

the “crystral structure like” pose (Figure 6). This destabilization is due to the decreasing hydrophobicity of the sugars of M-III and M-VI when placed in the hydrophobic pocket (Figure 7). Figure 7 shows that the distance of the M-VI deoxysugar to the center of L227, L94, and L83 is the largest among the three substrates. For M-IV, the average distance from mycinose to the center of the hydrophobic groups is only 3.4 Å.

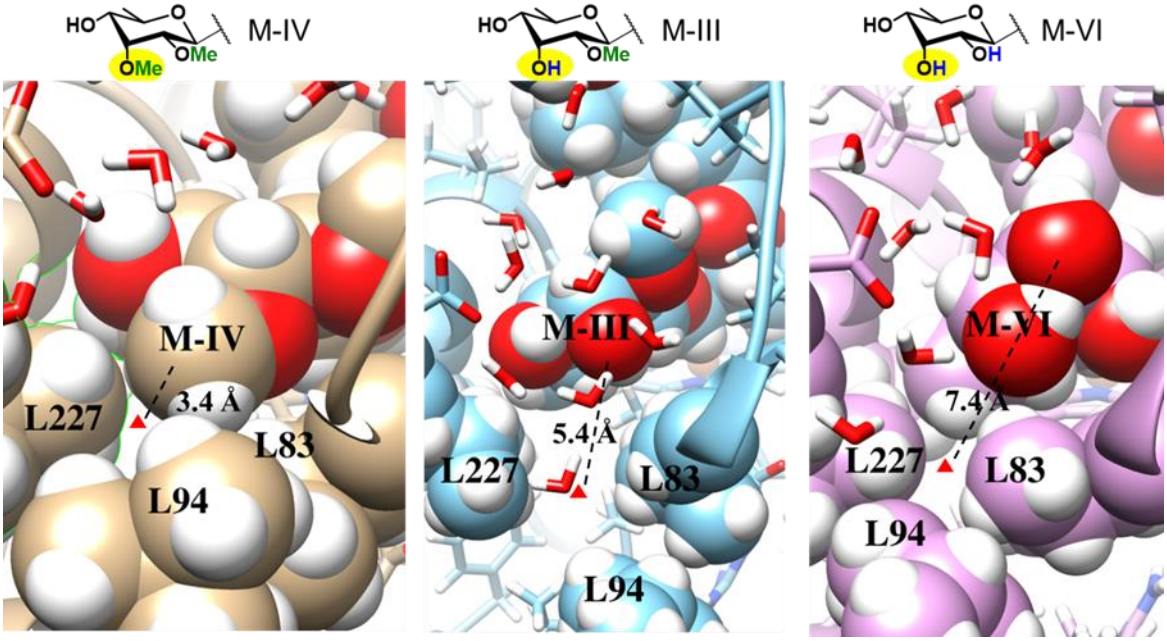


Figure 5.7. Deoxysugars of M-IV (left), M-III (middle), M-VI (right) bind in the hydrophobic pocket of L227, L94, L83. In chemdraw of each deoxysugar, the CH3 or OH that interact with the hydrophobic pocket is highlighted in yellow. Red triangle shown is the center point of a triangle created by the Cγ coordinates of L227, L83, L94.

The FG and BC loops contain E173 and E77, respectively, which were shown to form salt bridges with the desosamine of M-IV in the crystal structure. These loops also play a very important role in the substrate-MycG interactions. The two loops are “closed” when M-IV binds productively into MycG. The closed gate conformation is maintained by a new salt bridge between R75 and E173. However, for M-III and M-VI binding in MycG, this salt bridge interaction is not stable. The distance between the two residues are 12 Å and 20 Å for M-III and M-VI respectively (Figure 8). For M-IV, the two salt bridges with E173 and E77 are still observed in this closed

conformation, though salt bridge desosamine-E173 was broken after 100 ns. The simulation of M-III revealed that salt bridge desosamine-E77 persists throughout the entire MD simulation, but there is no salt-bridge interaction with E173. Both salt bridges are broken in the MD simulation of M-VI.

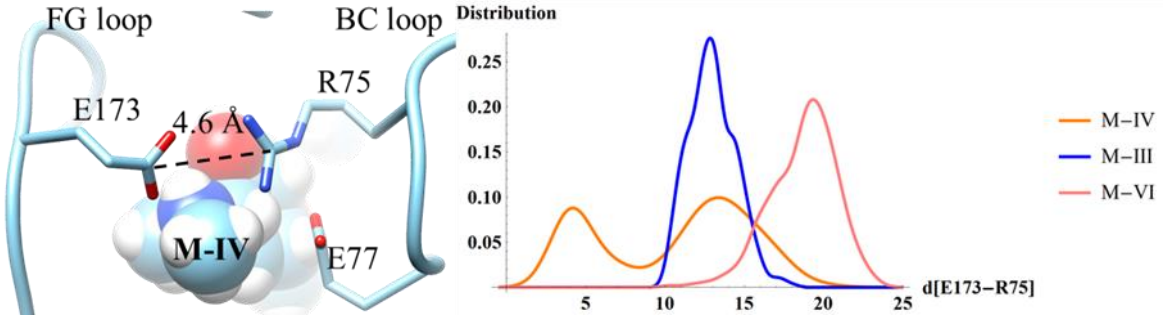


Figure 5.8. The distribution of the distance between E173 and R75 for M-IV(orange), M-III (blue) and M-VI (purple).

While the hydrophobic pocket and FG-BC gate are far from one another, they seem to both be linked to the substrate specificity of MycG. Putting all of the pieces together, we have determined how the enzyme is precisely designed for activity with M-IV. In the reactive pose discovered by constrained MD and metadynamics, the mycinose of M-IV binds in the hydrophobic pocket consisting of L227, L94, and L83 pocket. The binding shifts L227 residue away from L83 and results in movement of the I-helix proximal to the heme shown in Figure 5. Movement of this helix creates a more spacious pocket for macrolactone ring and allows C14 to get closer to iron-oxo compound. Moving of the whole macrolactone deeper into the binding pocket and facilitates the movement of the FG and BC loops closer together. The FG-BC gate closure furthermore keeps the salt bridges desosamine-E173, desosamine-E77 stable. This chain of interaction is pictured in Figure 9. When the hydrophobicity of the sugar at C14 is altered as in M-III and M-VI, the binding in the hydrophobic pocket is reduced. The substrate then sits higher in the binding pocket further from the iron-oxo and opening the FG-BC gate. In conclusion, the specificity of MycG toward M-

IV originates from the binding of less polar deoxysugar in a hydrophobic pocket, but is also a result of interactions in three different binding regions.

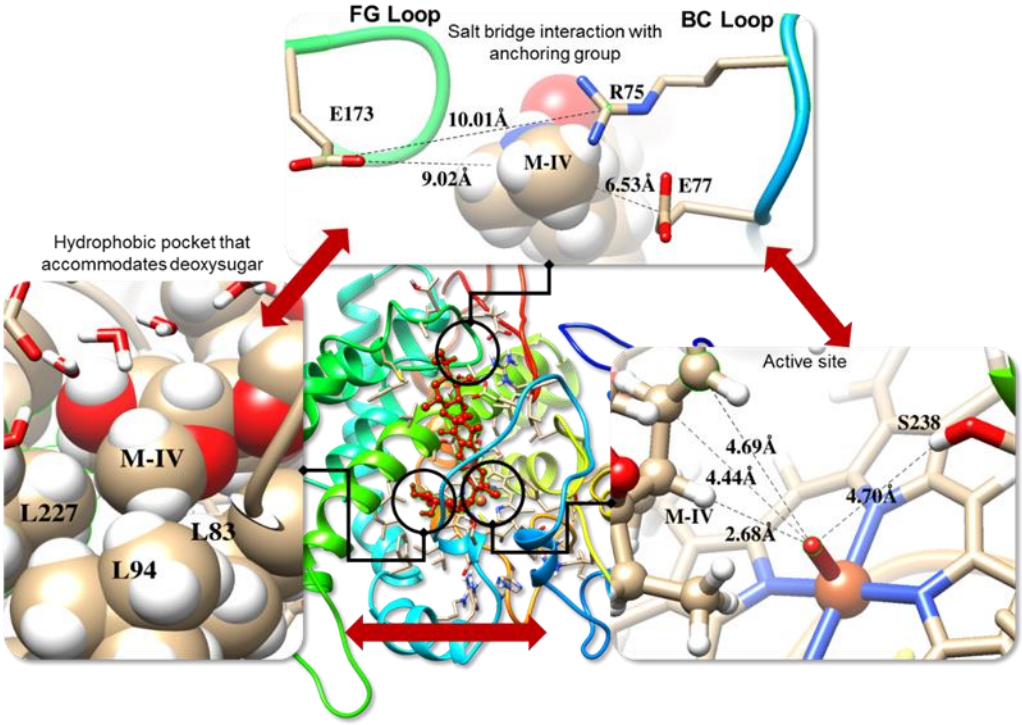


Figure 5.9. The connection between 3 different regions in the MycG-substrate interactions.

MD guided protein engineering. Several mutants were proposed based on the results discussed above to design an enzyme active towards M-III and M-VI. First, we studied mutations of the hydrophobic pocket which binds mycinose of M-IV. Figure 10 shows that mutations to this area do not appear to affect the distance profile of binding of M-VI. Mutant L227N and L227NL83NL94N show only moderate improvement in the HC14—OFe distance. L227NL83N is predicted to be worse than WT based on the 100 ns MD simulation. Next we studied if altering residues near the FG-BC gate would predict improved activity. In earlier simulations we noticed that E173 forms a deleterious salt bridge with R380 in the open form. We postulated that if R380 were removed, the gate would be less likely to adopt an open conformation and facilitate activity towards the non-natural substrates. To our delight, the R380E mutation maintains the HC14--OFe

distance below 3.0 Å for a majority of the simulation. We therefore predict improved activity toward M-VI if this mutation were made. These predictions indicate that binding in the hydrophobic binding pocket is complex and simply mutating leucines to hydrophilic residues does not always afford improved activity.

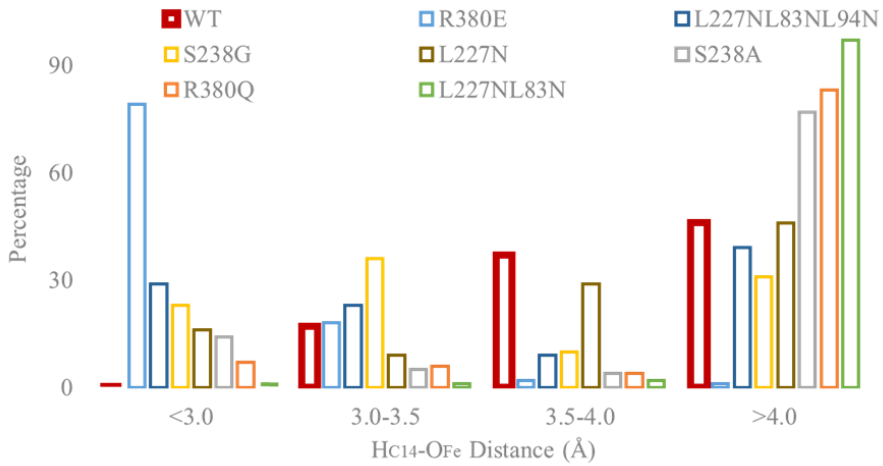


Figure 5.10. MD results of M-VI in WT (red bold) and proposed mutants, the distribution of distance HC14--OFe was extracted from the MD processes. R380E shows promising improvement by shifting the HC14--OFe distribution mainly below 3.0 Å.

5.5 Conclusions

We have used several computational methods to investigate the selectivity of the multi-functional P450 MycG. In the mycinamicin biosynthetic pathway to M-II, M-IV undergoes sequential hydroxylation to M-V then epoxidation to M-II. Density function theory calculations prove that the order of oxidation of M-IV is due to the need for stabilization on the C-H abstraction transition state. When the epoxide is installed first, leading to M-I, the presence of the epoxide has removed the allylic stabilization of M-IV and is inductively withdrawing. By using substrate analogues, we determined that the lack of resonance stabilization of the developing radical is the lead cause of the inactivity of MycG towards M-I. While MycG has activity towards M-IV, it has negligible activity toward M-III and M-VI, which only differ in the extent of methylation of the deoxysugar moiety. While a crystal structure of MycG with M-IV bound was obtained, it did not

show any reason why these 2 highly similar substrates are inactive. Metadynamics revealed a tighter, active binding pose which is more stable than the crystalized pose. This new pose places C14 closer to the iron-oxo and places the mycinose sugar of M-IV in a highly hydrophobic pocket. Molecular dynamics with M-III and M-VI bound show a decrease in stabilization of this active binding pose. The addition of hydrophilic groups to the mycinose sugar destabilizes the sugar in the hydrophobic pocket and pushes M-III and M-VI out of the active binding pose. Without the aid of these computational results, neither the substrate specificity nor order of reactivity could have been elucidated.

5.6 References

(1) Meesters, R. J. W.; Duisken, M.; Hollender, J. Cytochrome P450-Catalysed Arene-Epoxidation of the Bioactive Tea Tree Oil Ingredient p-Cymene: Indication for the Formation of a Reactive Allergenic Intermediate? Xenobiotica 2009, 39 (9), 663–671.

(2) Vaz, A. D. N.; McGinnity, D. F.; Coon, M. J. Epoxidation of Olefins by Cytochrome P450: Evidence from Site-Specific Mutagenesis for Hydroperoxo-Iron as an Electrophilic Oxidant. PNAS 1998, 95 (7), 3555–3560.

(3) Ortiz de Montellano, P. R. Hydrocarbon Hydroxylation by Cytochrome P450 Enzymes. Chem Rev 2010, 110 (2), 932.

(4) Lin, H.-C.; McMahon, T. C.; Patel, A.; Corsello, M.; Simon, A.; Xu, W.; Zhao, M.; Houk, K. N.; Garg, N. K.; Tang, Y. P450-Mediated Coupling of Indole Fragments To Forge Communesin and Unnatural Isomers. J. Am. Chem. Soc. 2016, 138 (12), 4002–4005.

(5) Usmani, K. A.; Karoly, E. D.; Hodgson, E.; Rose, R. L. In Vitro Sulfoxidation of Thioether Compounds by Human Cytochrome P450 and Flavin-Containing Monooxygenase Isoforms with Particular Reference to the CYP2C Subfamily. Drug Metab. Dispos. 2004, 32 (3), 333–339.

(6) Lee, H. S.; Park, E. J.; Ji, H. Y.; Kim, S. Y.; Im, G.-J.; Lee, S. M.; Jang, I. J. Identification of Cytochrome P450 Enzymes Responsible for N -Dealkylation of a New Oral Erectogenic, Mirodenafil. Xenobiotica 2008, 38 (1), 21–33.

(7) Anzai, Y.; Li, S.; Chaulagain, M. R.; Kinoshita, K.; Kato, F.; Montgomery, J.; Sherman, D. H. Functional Analysis of MycCI and MycG, Cytochrome P450 Enzymes Involved in Biosynthesis of Mycinamicin Macrolide Antibiotics. Chem Biol 2008, 15 (9), 950–959.

(8) Anzai, Y.; Tsukada, S.; Sakai, A.; Masuda, R.; Harada, C.; Domeki, A.; Li, S.; Kinoshita, K.; Sherman, D. H.; Kato, F. Function of Cytochrome P450 Enzymes MycCI and MycG in Micromonospora Griseorubida, a Producer of the Macrolide Antibiotic Mycinamicin. Antimicrob. Agents Chemother. 2012, 56 (7), 3648–3656.

(9) Tietz, D. R.; Podust, L. M.; Sherman, D. H.; Pochapsky, T. C. Solution Conformations and Dynamics of Substrate-Bound Cytochrome P450 MycG. Biochemistry 2017, 56 (21), 2701–2714.

(10) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. Gaussian 09, Revision B.01; Gaussian, Inc.: Wallingford CT, 2009.

(11) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; et al. Amber 14; University of California, San Francisco, 2014.

(12) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J. Chem. Theory Comput. 2013, 9 (9), 3878–3888.

(13) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. Computer Physics Communications 2013, 184 (2), 374–380.

(14) Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E. Quantum Mechanically Derived AMBER-Compatible Heme Parameters for Various States of the Cytochrome P450 Catalytic Cycle. J. Comput. Chem. 2012, 33 (2), 119–133.

(15) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. Journal of Molecular Graphics and Modelling 2006, 25 (2), 247–260.

(16) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. J. Comput. Chem. 2004, 25 (9), 1157–1174.

(17) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. J. Phys. Chem. 1993, 97 (40), 10269–10280.

(18) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. Proc Natl Acad Sci U S A 2002, 99 (20), 12562–12566.

(19) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. J. Comput. Chem. 2005, 26 (16), 1781–1802.

Chapter 6. Understanding Flavin Dependent Halogenase Selectivity via Molecular Dynamics Simulations of Laboratory-Evolved Halogenase Variants

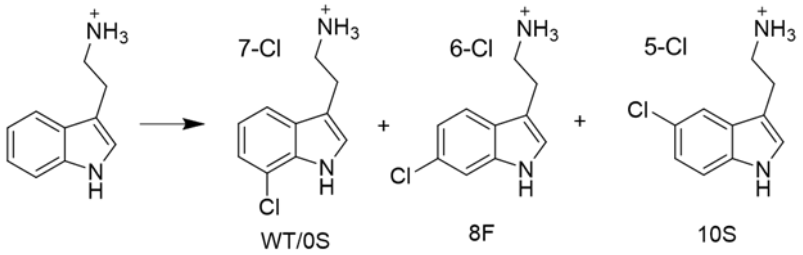
6.1 Abstract

The Lewis group reported the directed evolution of RebH, an flavin-dependent halogenase, to achieve highly regioselective 5-, 6-, or 7-chlorination of tryptamine. While the wild type enzyme catalyzes 7-chlorination of tryptamines, mutants favor either 5- or 6-chlorination. We applied molecular dynamics (MD) simulations in combination with DFT calculations to reveal how regioselectivity control is achieved by these enzymes. We have identified how mutations dictate the modes of substrate binding to achieve reactions at the 5-, 6-, or 7- positions.

6.2 Introduction

Directed evolution[1] can evolve enzymes to achieve highly active and selective enzyme catalysis. However, with many rounds of mutations and various combinations of residue changes, it is often difficult to identify which mutations trigger changes in selectivity or reactivity. The directed evolution process also does not provide insight about the mechanism behind how certain mutations enhance activity or change selectivity.

Scheme 6.1. Directed evolution results



RebH is a flavin-dependent halogenase that catalyzes chlorination of tryptophan and tryptamine.[2] The wildtype (WT) RebH chlorinates tryptophan or tryptamine at the 7-position of the indole moiety. Lewis et al. reported the directed evolution of RebH that gave rise to different RebH variants. Among these variants, 0S, 10S and 8F are of special interest.[3] 0S has a single

mutation, N470S, and it catalyzes 7-chlorination of tryptamine with higher activity than the WT. Further rounds of directed evolution led to mutant 10S variant that has 7 mutations and catalyzes 5-chlorination. The third mutant, 8F has 11 mutations and is a catalyst for 6-chlorination (Scheme 1).

While directed evolution resulted in RebH mutants that regioselectively chlorinate tryptamines at different positions of the indole ring, the mechanism behind the change in selectivity is unclear. We used molecular dynamics to show how the mutations in 0S, 10S and 8F influence the enzyme catalysis. We found that the binding modes of tryptamine to RebH are related to regioselectivity in 10S and 8F. Reverse mutation experiments were carried out in 10S and 8F to locate the key mutations responsible for the switch of selectivity. The reverse mutation results confirmed the key roles of mutations at residue 52 and 465.

6.3 Computational Methods

QM

All geometries were fully optimized using the B3LYP-D3 (BJ) functional with dispersion corrections and the triple-zeta 6-311+G(d,p) basis set in implicit pentyl ethanoate solvation (ε=4.7297) with the SMD implicit solvation model. This was used to account for the hydrophobic binding pocket in the enzyme. All optimized geometries were verified by frequency computations as minima (zero imaginary frequencies) or transition structures (a single imaginary frequency).

Docking

The RebH WT crystal structure (PDB: 2OA1)[4] was used for docking. The tryptophan cocrystallized in the RebH WT enzyme was removed before docking. Docking were carried out with Autodock Vina.[5] 0S, 10S, and 8F mutations were built based on the WT crystal structure by making the corresponding mutations with Chimera.[6] The receptors (WT, 0S, 10S and 8F) were

prepared for docking using Autodock Tools.[7] The same grid boxes of size 30×30×30Å3 were used to define the binding pocket for each RebH variant. Tryptamine was prepared as the docking ligand using Autodock Tools.[5] Nine docking conformations were obtained in the docking results. Two binding conformations relevant to the chlorination were identified in the docking results. The two binding conformations were further prepared for MD simulation of each RebH variant.

MD simulations

The FAD cofactor was modeled using the parameters from AMBER parameter database developed by Stuchebrukhov et al.[8] Simulations were performed using the GPU code (pmemd) of the AMBER 12 package.[9-10] Parameters for tryptamine substrate were generated within the antechamber module using the general AMBER force field (GAFF),[11] with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model.[12] Residue K79 was modified to be lysine chloramine $K79NH_2Cl^+$, and the parameters are generated using MCPB.py in the AMBER 12.[13] The charges were calculated according to the Merz-Singh-Kollman scheme, using the Gaussian 09 package.[14-16] The initial substrate-enzyme complexes were obtained from the docking results. Each protein was immersed in a pre-equilibrated truncated cuboid box of water with a 10 Å buffer of TIP3P water molecules using the leap module,[17] resulting in the addition of around 17600 solvent molecules. The systems were neutralized by addition of explicit counterions (Na^+ and Cl^-). All subsequent calculations were done using the widely tested Stony Brook modification of the Amber 99 force field (ff99sbildn).[18] A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on the tryptamine by a harmonic potential with a force constant of 500 kcal/(mol*Å)2. The second stage minimizes all the atoms in the simulation cell except those involved in the harmonic distance restraint. The systems were gently heated to 300

K, incrementing the temperatures by 5 K each step under constant-volume and periodic-boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between the hydrogen atoms was fixed. Long-range electrostatic effects were modelled using the particle-mesh-Ewald method.[19] An 8 Å cutoff was applied to Lennard-Jones and electrostatic interactions. Harmonic restraints of 30 kcal/mol were applied to the substrate, and the Andersen equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated for 2 ns with a 2 fs time step at a constant volume. Production trajectories were then run for an additional 500 ns under the same simulation conditions. Two trajectories were obtained from MD simulations for each RebH variant.

6.4 Results and Discussion

Quantum Mechanical Calculations

In a previous study, van Pee et al. reported that two residues in the RebH active site, K79/E357, are crucial in the chlorination of tryptophan.[20] The amine of K79 is chlorinated by HOCl generated by the FADH2, and the glutamate serves as a base to deprotonate and activate the ammonium of K79 for chlorination. We performed DFT calculations on a simple model system to assess the geometrical requirements for the chlorination step in the electrophilic aromatic substitution (Figure 1). The tryptamine is truncated to an indole, and the electrophile lysine-chloramine is truncated to an N-methyl-chlorammonium ion.

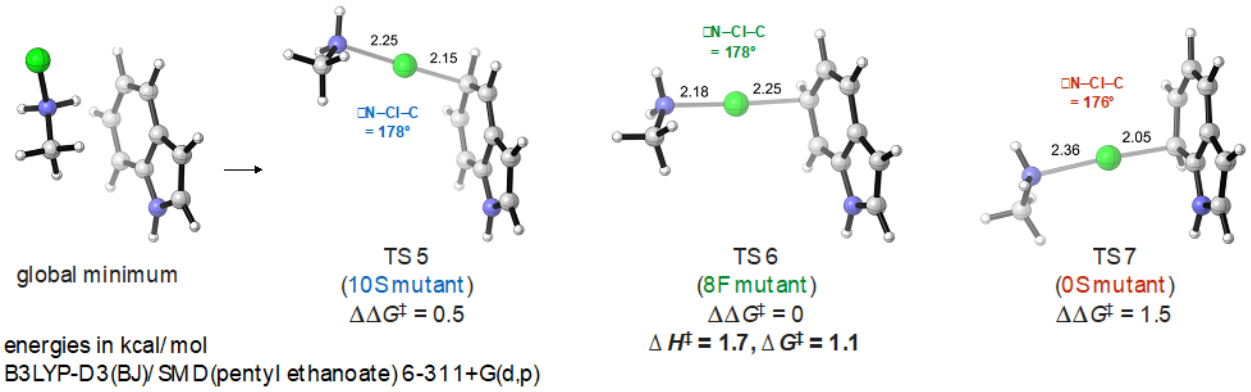


Figure 6.1. A QM model of the uncatalyzed electrophilic aromatic substitution between indole and N-methyl-chlorammonium.

Figure 1 shows the global minimum of the indole-chlorammonium complex and the transition structures (TSs 5, 6, and 7) that lead to the 5-, 6-, and 7-chloroindole products from left to right. TS 6 is the lowest-energy transition structure and has a free energy of activation of only 1.1 kcal/mol with respect to the complex. This suggests that the chlorination step is remarkably facile. In all three transition structures, the N–Cl–C angles are close to 180º. TS 5 and TS 7 are only 0.5 kcal/mol and 1.5 kcal/mol higher in energy than TS 6. Therefore, there is little intrinsic selectivity expected for such a reactive species. The orientation of the indole relative to the K79–NH_2Cl+ is more likely to control selectivity in the enzyme.

Binding Conformations of Tryptamine

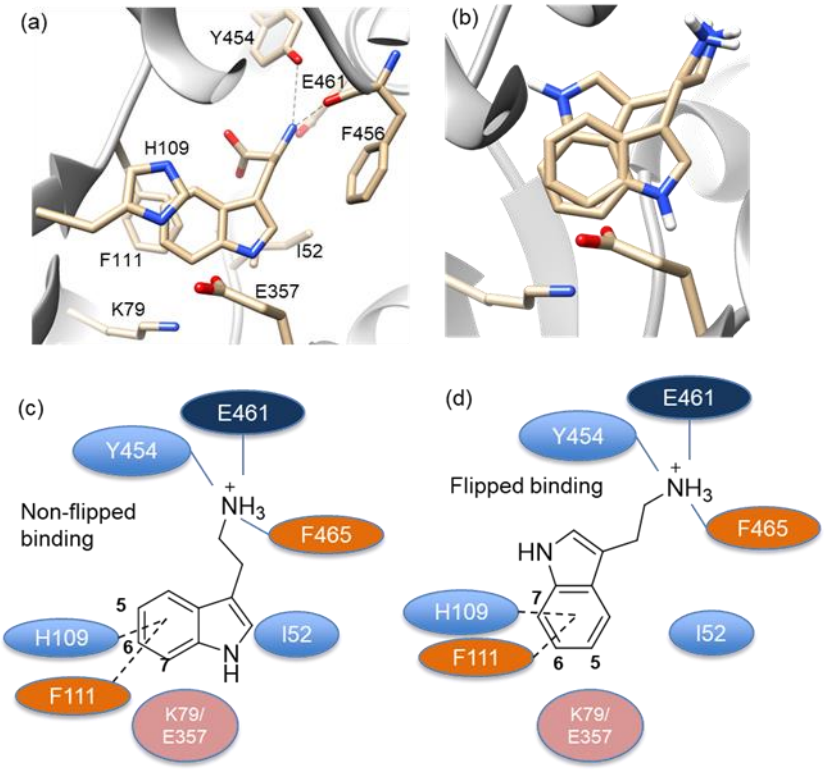


Figure 6.2. (a) Crystal structure of RebH with tryptophan bound in the binding pocket. (b) Two docking poses of tryptamine in RebH WT. Tryptamine can bind in the binding pocket in (c) non-flipped and (d) flipped binding modes.

Figure 2a shows the WT RebH crystal structure with tryptophan co-crystallized in the enzyme active site, with the important residues involved in the catalysis. For tryptamine, however, two binding modes (2b) were found to be favorable in docking. One is the same as the binding of tryptophan in the RebH crystal structure, with the nitrogen atom in the indole ring facing E357 (2c). The other binding mode is achieved by flipping the indole ring, which moves nitrogen atom away from E357 (2d). In the non-flipped binding mode, the ammonium group forms multiple stabilizing interactions with residues E461, Y454 and F465. Specifically, it forms a strong salt bridge with E461 at 3.5 Å and two hydrogen bonds with the hydroxyl oxygen on Y454 and the carbonyl group of the F465 backbone. These favorable interactions firmly anchor the flexible

linear chain of the tryptamine substrate. The indole ring is further stabilized by π-π interactions with the aromatic side chains of F111 and H109 in a sandwiched geometry.

The second binding mode is referred to as flipped binding mode shown in Figure 2d. The flipped binding conformation is achieved by rotating the tryptamine in the non-flipped binding conformation (2c) along one of the flexible C–C single bonds. The interaction between the ammonium moiety and the surrounding residues is identical to the non-flipped binding mode. The phenyl rings are in similar positions in for both binding conformations, which indicates that the π-π interactions with H109/F111 are also preserved in the flipped conformation. However, the E357-indole NH hydrogen bond is absent.

The two binding modes bring different sites on the phenyl ring close to K79/E357, the catalytic residues. In the nonflipped binding mode, we find that a hydrogen-bond interaction between E357 and the N–H group of the indole can orient the tryptamine in a way that the 7-position is in proximity to these two key residues. This allows chlorination at the C7 position, in agreement with the observed regioselectivity of the WT enzyme and the 0S mutant. However, this binding mode is unlikely for chlorination at C5 or C6 positions. The flipped binding mode places C5 and C6 of the indole ring close to K79/E357.

Molecular Dynamics Simulations

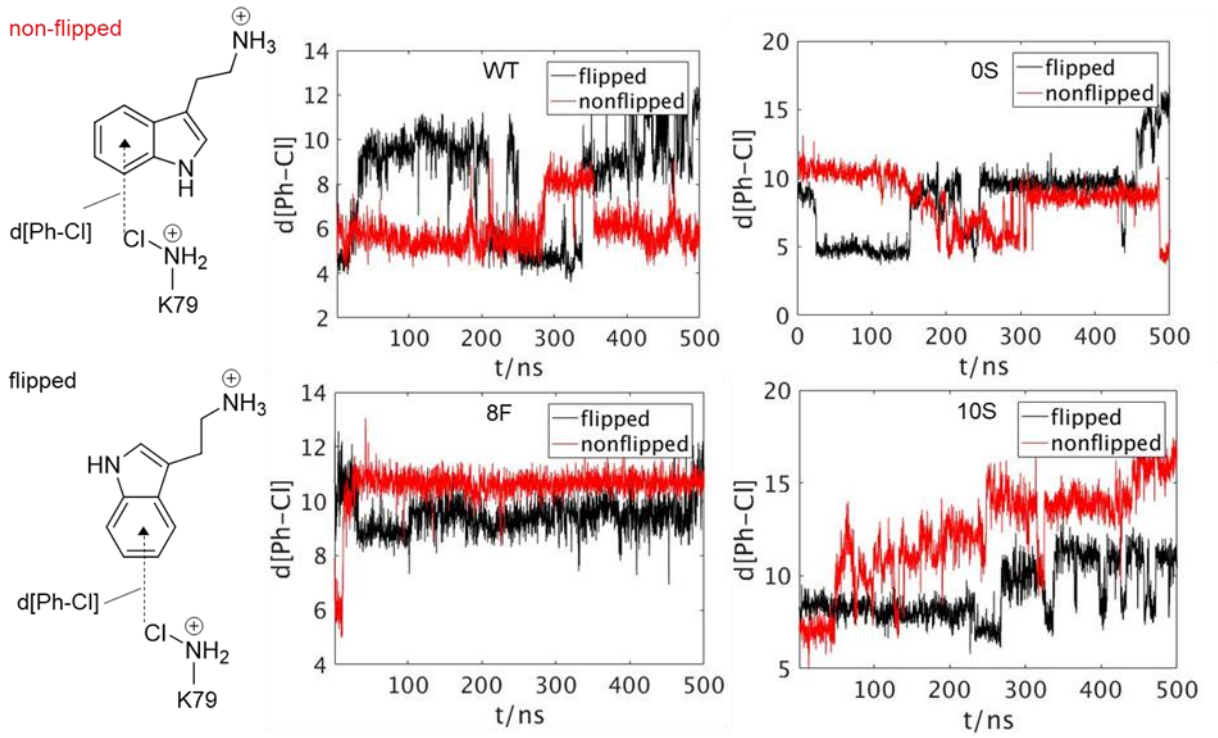


Figure 6.3. MD simulations of two binding modes for RebH WT and mutants. Distances between the phenyl center and the Cl atom in K79 were plotted for non-flipped (red) and flipped (black) docking poses.

Starting from the two binding conformations found in docking results, we carried out 500 ns MD simulations for RebH WT and the three mutants. The distances between the Cl atom in K79-NH2Cl+ and the phenyl center in the indole moiety in each snapshot were extracted from the MD trajectories. The phenyl center is defined as the geometrical center of the benzene ring. We tested whether the reactivity in the RebH-catalyzed chlorination is correlated to the distance between the phenyl center and the catalytic center. Figure 3 shows the distances for both binding conformations in all RebH variants. In WT and 0S mutant, the nonflipped binding conformation (red) places the indole ring closer to K79-NH2Cl+. In the flipped binding mode (black), the Ph-Cl distances largely fluctuate and climbs to 10-14 Å. On the other hand, 8F and 10S keep the tryptamine close to K79-NH2Cl+ in the flipped binding conformation (black), while the substrate is 10-12 Å away from K79-NH2Cl+ in the nonflipped binding mode (red).

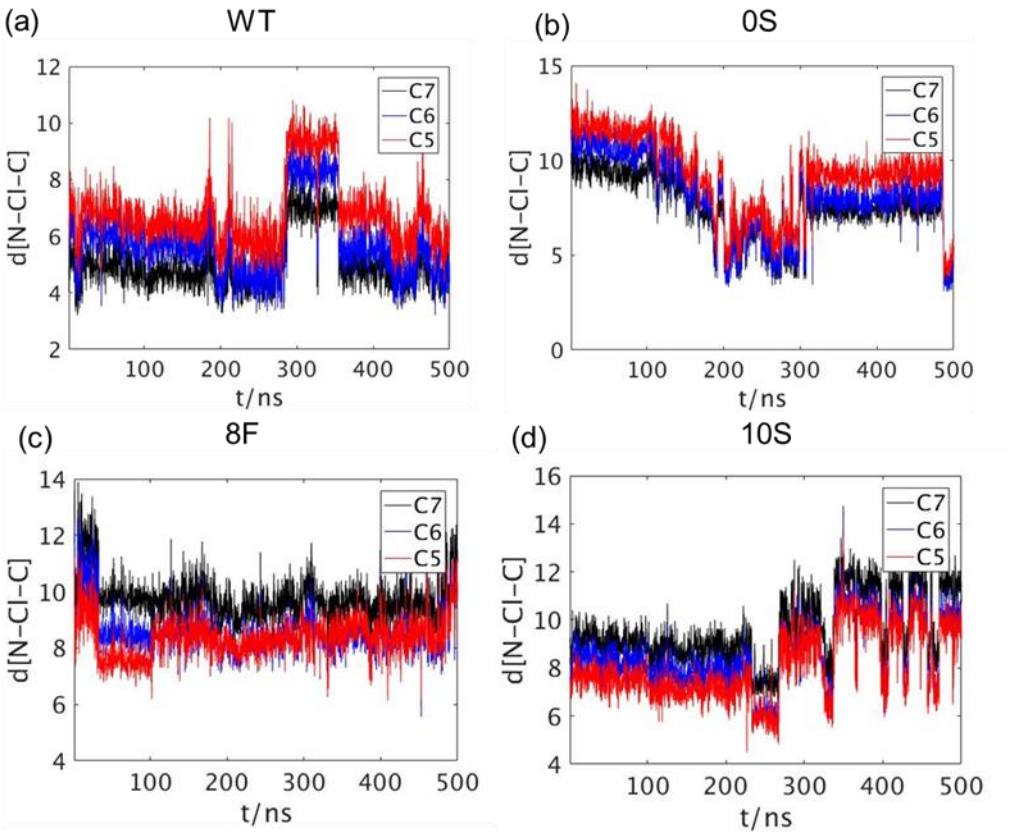


Figure 6.4. Plots of distance between C and Cl atom at K79. The distribution of the distances and angles are plotted for non-flipped binding mode in WT, 0S and flipped binding mode in 8F, 10S.

Table 6.1. Percent of MD simulation time for each carbon to be closest to Cl atom at K79

	WT	0S	8F	10S
C7	91%	95.04%	0%	0%
C6	9%	4.96%	47.04%	6.68%
C5	0%	0%	52.96%	93.28%

We further explored the nonflipped binding modes for WT and 0S. We plotted the distance from the Cl atom at K79 to the carbons (Figure 4). For the WT and 0S (Figure 4a, 4b), the C7 position (black line) of the substrate is closest to K79 in the non-flipped binding conformation. 91% of the MD snapshots have C7 closest to Cl atom at K79 (Table 1) in WT. Similarly, distance

from C7 to Cl atom at K79 is the minimum for 95.04% of the time in 0S. These results are in agreement with the experimentally observed 7-chlorination by WT and the 0S variant. For the 8F, the C5 positions (red dots) are closest to K79 from 0 ns to 200 ns in the flipped binding mode. After 200 ns, the distances C6-Cl and C5-Cl overlap largely. For 47.04% of the simulation time, the minimum distance is the distance from C6 to K79. In 10S, K79 is closer to the C5 position (red dots) in the indole ring all the time, which leads to chlorination at the C5 position (Figure 4d). For 10S, most of the time (93.28%) C5 is closest to K79.

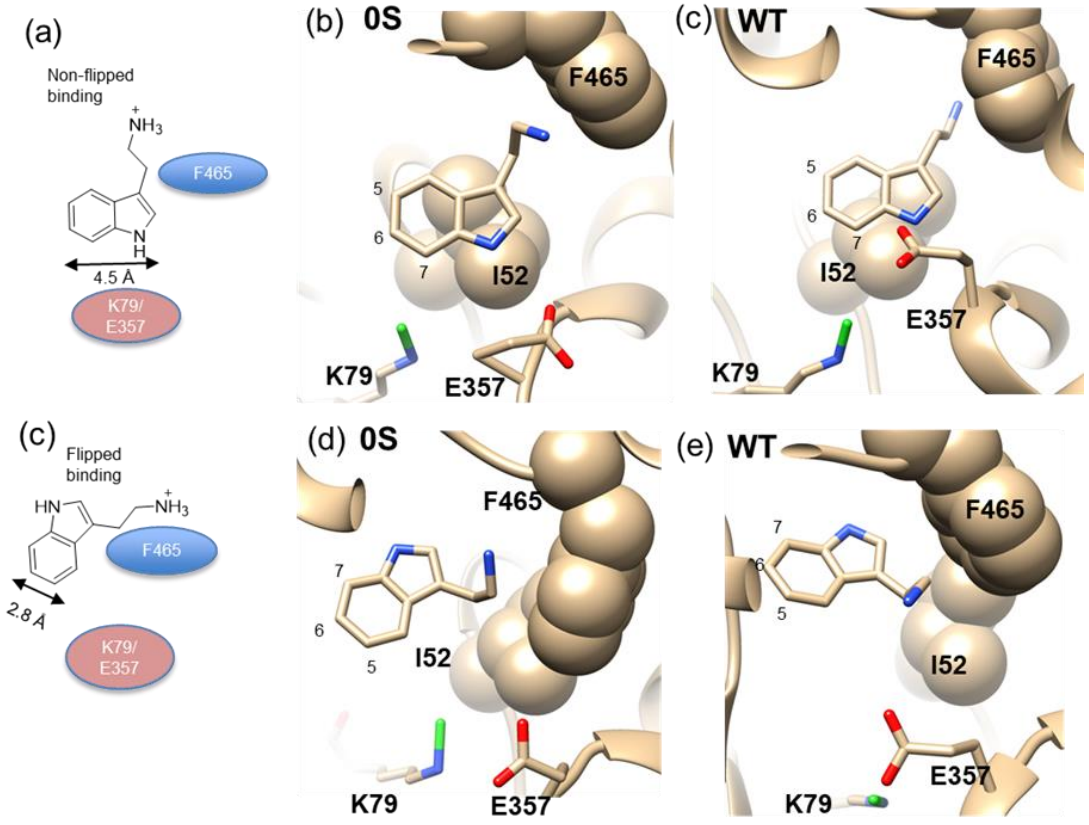


Figure 6.5. Representative snapshots from MD simulations: (b) non-flipped binding conformation in 0S; (c) non-flipped binding conformation in WT; (d) flipped binding conformation in 0S; (e) flipped binding conformation in WT.

Figure 5 shows representative snapshots from MD simulations of WT and 0S. We compared the geometry of the tryptamine, especially the indole ring, in the two binding modes.

Two specific residues close to the tryptamine, F465 and I52 are in the van der Waals contact with the indole ring in MD simulations. In the nonflipped binding mode (Figure 5b, 5c), reside I52 is below the tryptamine, and van der Waals contact stabilizes the nonflipped binding conformation. In the flipped binding conformation (Figure 5d, 5e), I52 is in van der Waals contact with F465, and the side chain of tryptamine is in contact with F465. In this arrangement, tryptamine cannot get close to the catalytic residues K79 and E357.

Figure 6 shows representative snapshots from MD simulations of 10S and 8F. Residue I52 is mutated to histidine and methionine in 10S and 8F, respectively. F465 has been mutated to cysteine and leucine, respectively. Both histidine and methionine are bulkier than the original isoleucine 52. As a result, residue 52 is in the same plane as the indole ring of tryptamine for both 10S and 8F. Moreover, the original bulky F465 is mutated to smaller residues L465 or C465, which results in more space for the aryl side chain in the flipped binding mode (Figure 6d, 6e). In the nonflipped binding mode (Figure 6b, 6c), H52/M52 are in van der Waals contact with tryptamine from one side of the indole ring. The steric clash with residue 52 prevents tryptamine getting close to K79/E357 in the nonflipped binding mode. The flipped binding conformation places the phenyl side of the indole ring close to residue 52. The phenyl side fits in the binding pocket without suffering steric clash with residue 52.

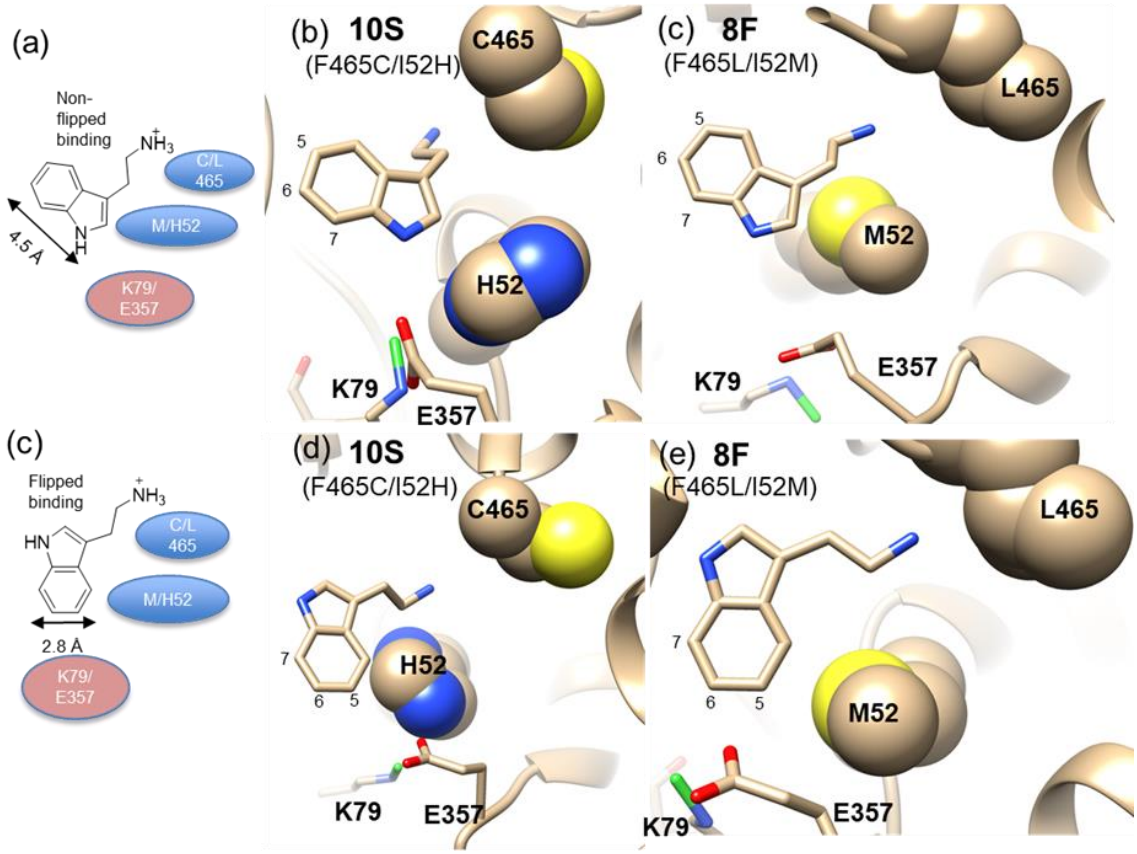


Figure 6.6. Representative snapshot of MD simulations: (b) non-flipped binding conformation in 10S; (c) non-flipped binding conformation in 8F; (d) flipped binding conformation in 10S; (e) flipped binding conformation in 8F.

We compared the MD results of 8F and 10S. The difference between configurations that can lead to 5- and 6-chlorination is small. The C5 and C6 positions on the indole are exposed to the active K79/E357. Experimentally 8F prefers 6-chlorination while 10S tends to attack the tryptamine at C5 position. The binding conformations of tryptamine in the active sites of 8F and 10S are similar. The difference comes from the orientation of E357 (Figure 8). In 8F, E357 is pointing more to the 6-position. In 10S, E357 is closer to the 5-position. The difference in E357 orientation comes from its different interactions with residue 52 in 8F and 10S. In 8F, residue 52 was mutated to methionine with a bulky thioether side chain. Consequently, E357 moves away from M52 due to steric repulsion. In 10S, residue 52 is mutated to histidine, which contains a

hydrogen-bond donor on the imidazole side chain. This favorable electrostatic interaction steers E357 towards H52

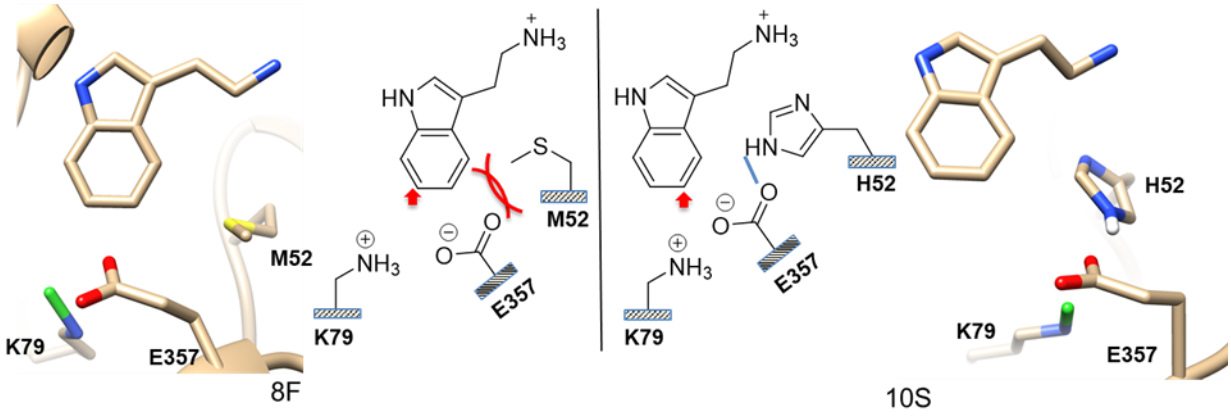


Figure 6.7. Interactions of tryptamine with the three important residues in 8F and 10S. Mutants 8F and 10S use the same strategy to make the substrate adapt the flipped binding mode. The major difference between 10S and 8F is residue 52. In 8F, M52 residue is a bulky hydrophobic residue, which has repulsion with E357. In 10S, H52 residue is a hydrogen-bond donor, which attracts E357.

Reverse Mutations

Reversed mutations were carried out for 10S and 8F to identify the key mutations. In 10S, reverse mutations Q509R, F380L and R494Q have little to no effect to the reactivity or selectivity. S470N and S470A reduce the yield of 5-chlorination products largely, but they remain 5-selective. C465F almost kills the reactivity. However, the 5-chlorotryptamine product is still the major product. Only reverse mutation H52I switched the selectivity back to 7-chlorination. Similarly, in 8F, only reverse mutation M52I shows switch of regioselectivity to 7-chlorination. L465F is still 6-selective but with much lower yields. The reverse mutations show that the mutations at residue 52 are the cause of regioselectivity change, and mutations at residue 465 are also essential to achieve high yield in 5- and 6- chlorination.

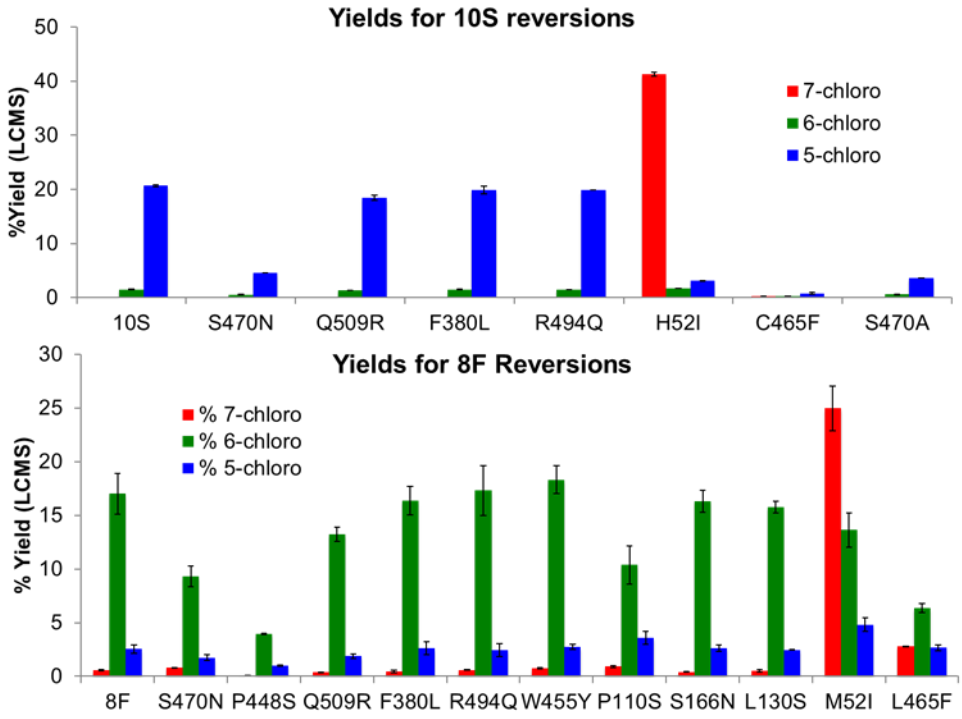


Figure 6.8. Reverse mutations starting from 8F and 10S show that H52I and M52I reverse the selectivity back to 7-chlorination. Meanwhile reverse mutation C465F and L465F result in dramatic decrease in yield of chlorination products.

6.5 Conclusions

We studied the dynamics of the RebH enzyme and three variants (0S, 8F, and 10S) obtained from directed evolution. QM calculations showed that the activation barriers for chlorination at each of the three positions are only 1¬–3 kcal/mol, and the regioselectivity is dictated by how the tryptamine is bound in the active sites. For each case, a pair of key mutations was identified to be responsible for the change in regioselectivity. The combination of I52M/F465L or I52H/F465C reshapes the RebH binding pocket. This change leads to a change in preference for the binding conformation of tryptamine, which results in chlorination at the C6 position for 8F and C5 position for the 10S mutant.

6.6 References

1. F. H. Arnold, Design by Directed Evolution. Acc. Chem. Res. 31, 125–131 (1998).

2. E. Yeh, S. Garneau, C. T. Walsh, Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. PNAS. 102, 3960–3965 (2005).

3. M. C. Andorfer, H. June Park, J. Vergara-Coll, J. C. Lewis, Directed evolution of RebH for catalyst-controlled halogenation of indole C–H bonds. Chemical Science. 7, 3720–3729 (2016).

4. E. Bitto et al., The structure of flavin-dependent tryptophan 7-halogenase RebH. Proteins. 70, 289–293 (2008).

5. O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry 31 (2010) 455-461

6. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

7. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. J. Computational Chemistry 2009, 16: 2785-91

8. J. Antony, D. M. Medvedev, A. A. Stuchebrukhov, Theoretical Study of Electron Transfer between the Photolyase Catalytic Cofactor FADH- and DNA Thymine Dimer. J. Am. Chem. Soc. 122, 1057–1065 (2000).

9. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J. Chem. Theory Comput. 9, 3878–3888 (2013).

10. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2012), AMBER 12, University of California, San Francisco.

11. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of a general amber force field. J. Comput. Chem. 25, 1157–1174 (2004).

12. C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J. Phys. Chem. 97, 10269–10280 (1993).

13. P. Li, K. M. Merz, MCPB.py: A Python Based Metal Center Parameter Builder. J. Chem. Inf. Model. 56, 599–604 (2016).

14. U. C. Singh, P. A. Kollman, An approach to computing electrostatic charges for molecules. J. Comput. Chem. 5, 129–145 (1984).

15. B. H. Besler, K. M. Merz, P. A. Kollman, Atomic charges derived from semiempirical methods. J. Comput. Chem. 11, 431–439 (1990).

16. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota,

R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09 (Gaussian, Inc., Wallingford CT, 2009).

17. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. 79, 926–935 (1983).

18. K. Lindorff-Larsen et al., Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 78, 1950–1958 (2010).

19. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics. 98, 10089–10092 (1993).

20. S. Flecks et al., New Insights into the Mechanism of Enzymatic Chlorination of Tryptophan. Angewandte Chemie International Edition. 47, 9533–9536 (2008).

Chapter 7. Crystal Fluidity Reflected by Fast Rotational Motion at the Core, Branches, and Peripheral Aromatic Groups of a Dendrimeric Molecular Rotor

7.1 Abstract

Low packing densities are key structural features of amphidynamic crystals built with static and mobile components. Here we report a loosely packed crystal of dendrimeric rotor 2 and the fast dynamics of all its aromatic groups, both resulting from the hyperbranched structure of the molecule. Compound 2 was synthesized with a convergent strategy to construct a central phenylene core with stators consisting of two layers of triarylmethyl groups. Single crystal X-ray diffraction analysis confirmed a low-density packing structure consisting of one molecule of 2 and approximately eight solvent molecules per unit cell. Three isotopologues of 2 were synthesized to study the motion of each segment of the molecule in the solid state using variable temperature quadrupolar echo 2H NMR spectroscopy. Line shape analysis of the spectra reveals that the central phenylene, the six branch phenylenes, and the 18 periphery phenyls all display megahertz rotational dynamics in the crystals at ambient temperature. Arrhenius analysis of the data gives similar activation energies and preexponential factors for different parts of the structure. The observed pre-exponential factors are 4−6 orders of magnitude greater than those of elementary site-exchange processes, indicating that the dynamics are not dictated by static energetic potentials. Instead, the activation energies for rotations in the crystals of 2 are controlled by temperature dependent local structural fluctuations and crystal fluidity.

7.2 Introduction

While exploring the design of structures capable of supporting the functions of molecular machines,[1,2] our group has proposed the use of amphidynamic crystals featuring molecular gyroscopes and other molecular rotors with static frames and dynamic components.[3−5] Compound

1 is one of the simplest molecular gyroscopes with an open topology that we have analyzed in some detail [5f, g] (Figure 1). It consists of a phenylene rotator at the center linked by a double alkyne axle to a stator comprised of two triphenylmethyl (trityl) groups. The dumbbell shape of 1 makes it difficult for it to pack efficiently and generates a lowdensity region in the middle of the structure that, along with local volume fluctuations,4e allows for rotation of the central phenylene. Using variable temperature (VT) solid-state quadrupolar echo ^{2}H NMR spectroscopy (SS ^{2}H NMR), cross-polarization magic angle spinning (CPMAS) ^{13}C NMR spectroscopy, and computational analysis, we showed that the trityl stators engaged in a relatively tight 6-fold edge-to-face phenyl embrace are static, while the central rotator experiences a thermally activated 180° site exchange with a frequency of ca. 15 kHz at 297 K.[5f] More recently, we and others have described new examples of amphidynamic crystals with a range of internal dynamics.[4,6] Some of them have promising applications as novel materials with dielectric[4f,6a,d] and ferroelectric[6g,i,j] functions.

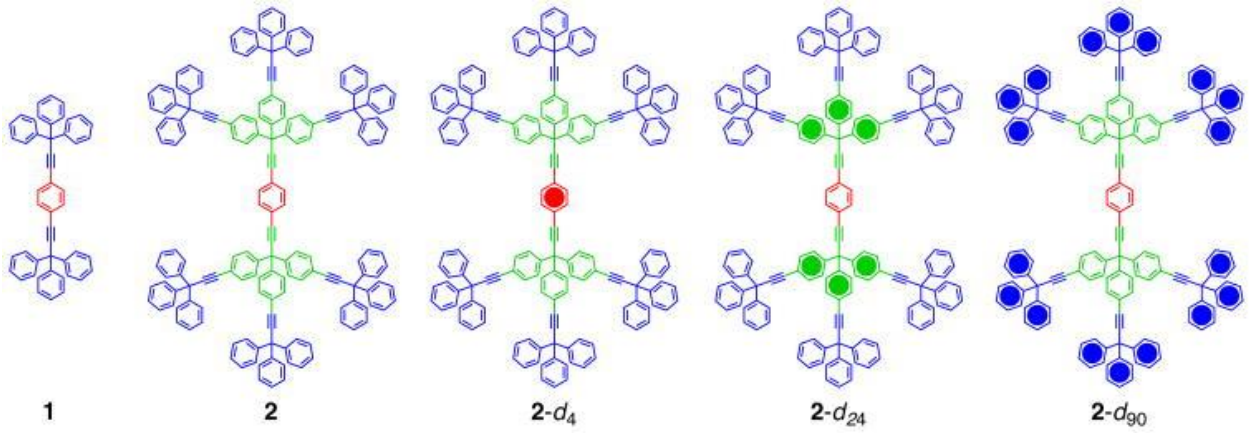


Figure 7.1. Structures of molecular rotor 1, dendrimeric molecular rotor 2, and its isotopologues 2-d4, 2-d24 and 2-d90 with the filled circles indicating perdeuterated phenylene and phenyl groups.

Considering strategies to increase the size of molecular gyroscopes to the range of biomolecular machines,[7] we decided to explore strategies to expand the architecture of 1 in a selfsimilar, radially growing manner. We viewed compound 2 as a promising target to obtain a well-ordered amphidynamic crystal based on its highly symmetric, shape-persistent,

hyperbranched structure, and its many shape-conserving conformational degrees of freedom. We report here the synthesis, structure, and dynamic processes in 2, a unique type of amphidynamic crystal. Crystals of this high molecular weight organic substance were obtained only from unconventional solvent mixtures. The crystal structure of 2 confirms the formation of a low-density crystal with a large amount of solvent molecules and intermolecular contacts reminiscent of a protein crystal. A close look at the structure suggested that motion might not be restricted to the central phenylene rotator (shown in red in Figure 1) but might extend to the phenylene branches (shown in green) and into the peripheral phenyl groups (shown in blue). With that in mind and taking advantage of a very efficient modular synthesis, we prepared deuterated isotopologues 2-d4, 2-d24, and 2-d90 to analyze their rotational dynamics. SS ^{2}H NMR measurements described here reveal ambient temperature rotational motion in the megahertz regime, indicating a crystal structure where every aromatic group displays fast rotational dynamics, whether it is the central phenylene rotator, phenylene groups in the trityl branches, or phenyl groups in the peripheral trityl groups. Crystals of dendritic rotor 2 represent a new type of amphidynamic crystals in which the frame of reference is solely determined by the C−C single bond framework and the quaternary carbons, which retain their equilibrium positions in the lattice while everything else rotates. Rotational dynamics as a function of temperature indicate that this unusually mobile environment is characterized by relatively high activation energies (Ea ≈ 15 kcal/mol) and very large preexponential factors (A ≈ 1018 s^{-1}). In a structure where every aromatic group undergoes thermally activated rotations, an unusually large pre-exponential factor suggests that those dynamic processes are influenced by changes in the internal fluidity of the crystal lattice, rather than a static potential energy hypersurface of a simple site-exchange process.

7.3 Computational Methods

Molecular Dynamics Simulation and Modes of Rotation in Vacuum

Molecular dynamics simulation of rotor 2 was performed using Amber 14.4 The parameters were generated with the antechamber module using the general Amber force field (GAFF), and the atomic partial charges were assigned using the AM1-BCC method. In the simulation, the structure of 2 was first minimized in vacuum with no solvent model. The system then was heated from 0 to 300 K under constant pressure of 1 atm in 2 ns before it was then equilibrated for 0.5 ns with 1 fs time step at a constant volume. Production trajectories were then run for additional 1 ns under the same simulation conditions. The dihedral angle change of each phenylene groups in one branch trityl unit was tracked and shown in Figure S47. Although much less frequent compared to the correlated motions, isolated rotation of one aromatic ring is indeed possible even in vacuum.

Estimated Free Energy Change of Rotation

The activation energies of branch phenylenes rotation in vacuum were studied using umbrella sampling. Potentials of mean force (PMF) were constructed using Weighted Histogram Analysis Method (WHAM) with the dihedral angles extracted every 50 fs.5 At each window, 1ns MD was performed with harmonic potential applied on the dihedral at 300 K under 1 atm pressure. When one phenylene explored 180-degree dihedral angle change from the lowest-energy-state with a 5-degree step width, the other two phenylene rings was allowed to relax to mimic correlated rotation (Figure S48, red circles). Umbrella constraint of 200 kcal/(mol•Å2) was applied to the chosen phenylene ring. The free energy change of rotation was found to be less than 1 kcal/mol in vacuum. Since the independent rotation of one phenylene is a higher energy process in vacuum, we could only estimate the activation energy by restraining a second phenylene while the chosen

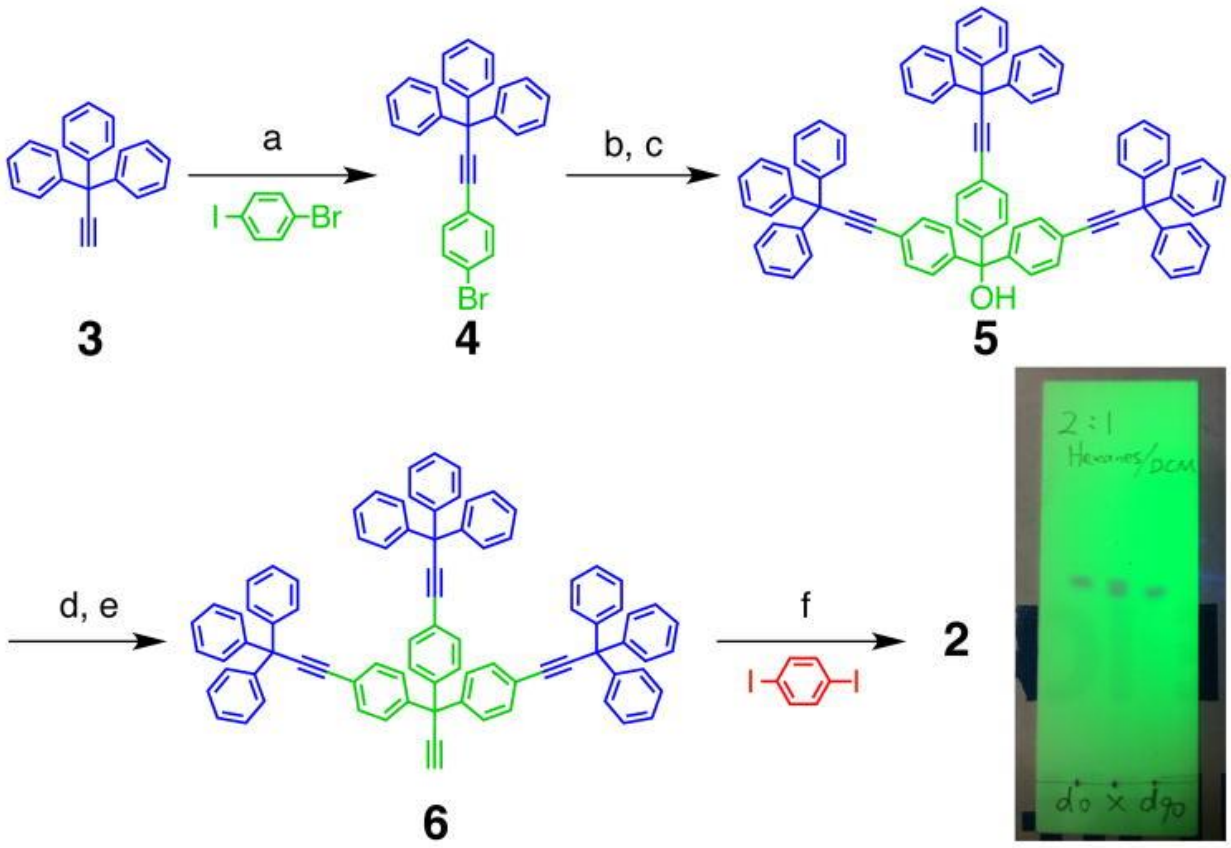
phenylene group was experiencing the 180-degree dihedral angle change as described above. The resulting energy diagram matched well with a 2-fold rotation model with activation energy of ca. 5 kcal/mol. If we restrain both the second and the third phenylenes, the energetic change would reflect the activation barrier to rotate one phenylene with the other two being completely static. The activation energy of this process in vacuum (ca. 19 kcal/mol) is higher than the experimental data (no more than 15 kcal/mol).

7.4 Results and Discussion

7.4.1 Synthesis and Characterization

Dendrimeric molecular rotor 2 was synthesized by taking advantage of the convergent strategy illustrated in Scheme 1. Sonagashira reaction between 3,3,3-triphenylpropyne 35g and 4-bromo-iodobenzene occurred selectively at the iodine site to give aryl bromide 4. Treatment of 4 with n-butyllithium provided an aryllithium species, which was trapped with diethyl carbonate to give triarylmethanol 5. When refluxed in a mixture of acetyl chloride and toluene, triarylmethanol 5 was converted to a triarylmethyl chloride intermediate, which was further reacted with ethynyl magnesium bromide to yield triarylpropyne 6. Although palladium catalyzed coupling reactions of two large components are generally difficult,8 we were able to achieve a double Sonagashira coupling reaction between terminal alkyne 6 and 1,4-diiodobenzene. Under optimized reaction conditions, rotor 2 could be obtained in 70% isolated yield and the homocoupling side reaction of 6 (not shown) was suppressed.

Scheme 7.1. Reagents and conditions: a) $PdCl_2(PPh_3)_2$, CuI, THF, i-Pr_2NH, reflux, 91 %; b) n-BuLi, THF; then c) diethyl carbonate, −78 °C to r.t., 85 %; d) AcCl, Tol, reflux; e) THF, Tol, reflux, 63 % over 2 steps; f) $Pd(PPh_3)_4$, CuI, NEt_3, PPh_3, DMF, 70 °C, 70 %. TLC plate illustrating the separation of 2 and 2-d90 in the middle lane.



Isotopologues 2-d4, 2-d24, and 2-d90 were prepared employing similar reactions with deuterated starting materials, 1,4dibromobenzene-d4 and 3-d15,9 to establish the desired isotopic labels. For details of the synthesis, please see the Supporting Information. Rotors 2, 2-d4, 2-d24, and 2-d90 were all fully characterized by solution ^{1}H and ^{13}C NMR, IR, and MALDITOF MS. Splitting of corresponding carbon signals could be observed in the ^{13}C NMR for rotors 2-d4, 2-d24, and 2-d90 due to the C−D couplings (J = 22−25 Hz). The intensities of C−H stretching peaks (3000−3100 cm−1) and C−H out-of-plane bending peaks (650−850 cm−1) in IR of isotopologues 2-d24 and 2-d90 were significantly reduced resulting from the deuterium substitutions. To our surprise, the retention value (Rf) of isotopologue 2-d90 (0.47) on a thin-layer chromatography (TLC) plate is different from those of rotors 2, 2-d4, and 2-d24 (0.49). This is a rather rare example of TLC separation of isotopologues,10 and it could be explained by a reduced hydrodynamic radius of the molecule due to the collective effects of polydeuteration at the periphery.11

7.4.2 Crystallization and X-ray Structure of Dendrimeric Molecular Rotor 2.

Crystallization of large organic molecules has always been a daunting challenge. To date, there are only a handful of crystal structures of compounds with molecular weight over 2000 Da reported in literature.[8b,12] Fortunately, we obtained good quality crystals from warm supersaturated solutions of 2 and its isotopologues in a nonconventional solvent mixture containing 2,4,6-trimethylpyridine and 2,2,4trimethylpentane.13 X-ray diffraction data was collected at 100 K, and the crystal structure was solved in the triclinic space group $P\bar{1}$. The asymmetric unit consists of one-half of the rotor molecule together with about four solvent molecules, most of them being 2,4,6-trimethylpyridine. Since some of the solvent molecules are highly disordered and could not be modeled accurately, they were removed using the SQUEEZE algorithm.[14] As shown in Figure 2, the three C−Ph bond vectors of the inner trityl groups on the sides of the structure adopt anti conformations, as is commonly seen in crystals of molecular gyroscopes of this type.5 A similar disposition can be observed for the peripheral trityls. Interestingly, all four trityl groups in the one-half of each rotor (Figure 2, left) have P or PPP helicity, while the other four related by an inversion center have M or MMM helicity. The inner alkyne axles display a slight deviation from linearity as measured by an angle of 175° measured from the ipso carbon of the central phenylene rotator to the center of the alkyne bond to the trityl quaternary carbon. Further analysis of the crystal structure shows the packing coefficient is only 0.46 without including the solvent molecules,15 which is very low compared with the typical range (from 0.64 to 0.77) for organic molecules.16 As a result, very few rotor−rotor close contacts were observed, mostly describable in terms of edge-to-face aromatic interactions. For example, the two faces of the central phenylene (C in Figure 3) are in close proximity to the edges of trityl branch phenylenes from two neighbors (B--C--B), and its two edges are directed toward the face of peripheral phenyl rings from neighboring

molecules (P--C--P). Each of the branch phenylenes has only one close contact either with the central phenylene (B--C) or with one of the peripheral phenyl rings (not shown). The peripheral phenyl rings have less than two close contacts on average with other phenyl/phenylene rings. These observations support our hypothesis that the hyperbranched structure would lead to fewer interactions in the crystal, which is imperative to the realization of fast dynamics. It should be noted that while solvent molecules are likely to be highly mobile, they also play an important role supporting the crystal structure, just like water molecules in protein crystals.[17] In fact, while attempting to obtain a crystal structure at 200 K, we noticed that all solvent molecules were disordered and none of them could be properly refined.

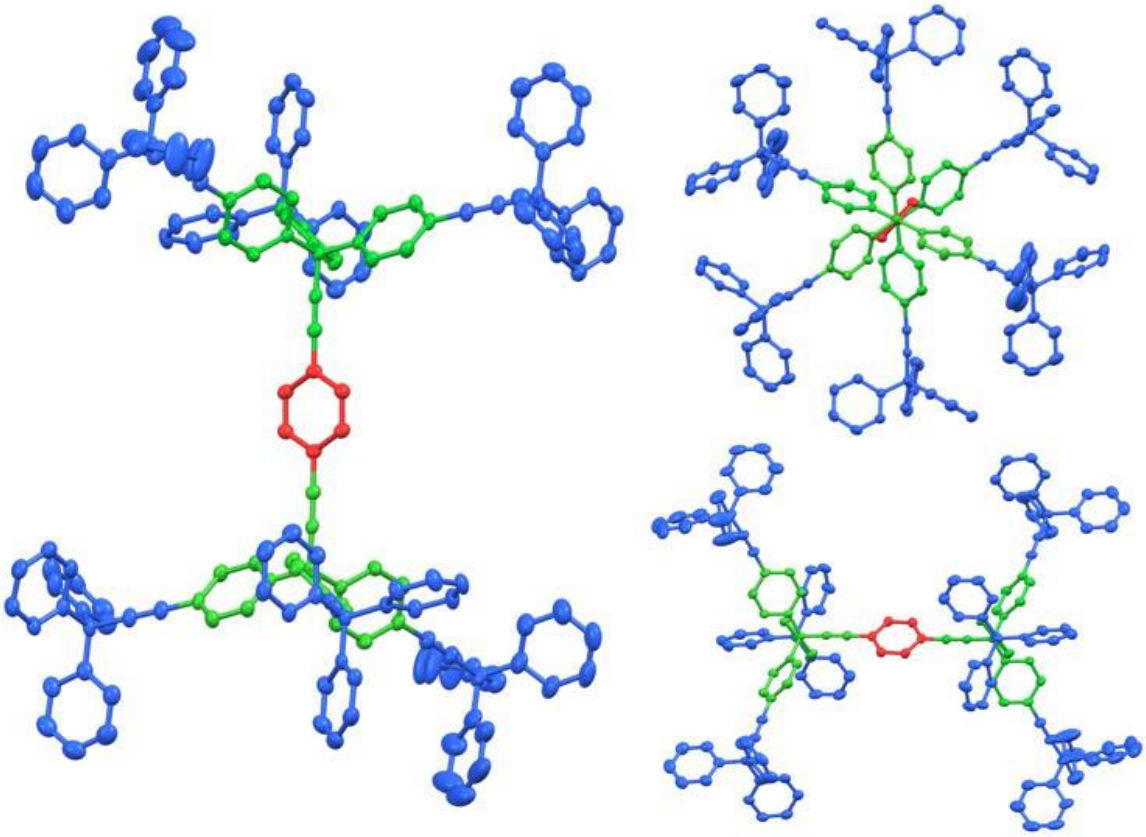


Figure 7.2. (Left) Crystal structure of dendritic rotor 2 with thermal ellipsoids showing 50% probability. Solvent molecules and hydrogen atoms are omitted for clarity. (Top right) View down the principal molecular axis and (bottom right) side view of 2 showing the anti conformations, respectively, between branch (green inner) trityls and the branch (green) and the peripheral (blue) trityls.

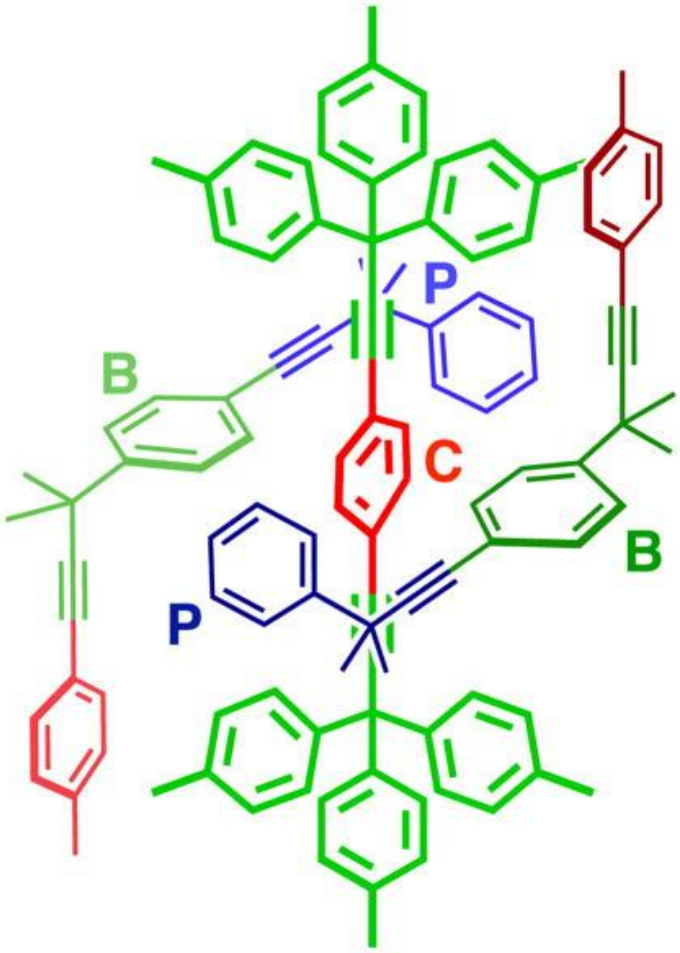


Figure 7.3. Schematic representation of the packing environment of the central phenylene (C) engaged in edge-to-face interactions acting as an acceptor with a branch phenylene (B) and as a donor with a peripheral (P) phenyl group.

7.4.3 Variable temperature solid state ^{2}H NMR experiments

^{2}H NMR spectroscopy is a widely used technique to analyze the internal dynamics of deuterium-enriched groups in the 103−108 Hz regime in the solid state.[18] The method relies on static sample measurements acquired with the quadrupolar spin echo pulse sequence and is based on the changes in line shape that result from the dynamic narrowing of the broad powder pattern in the static spectra. Since the line shape is sensitive to both the trajectory and frequency of the site exchange dynamics, simulation of the experimental spectra usually provides sufficient information to characterize motions in solids in great detail.[19] The ^{2}H NMR spin−echo experiments

were performed on crystalline and amorphous samples of 2-d4, 2-d24, and 2-d90 to explore the dynamics of the central phenylene, branch phenylenes, and peripheral phenyls selectively.[20] Spectra with good signal-to-noise ratios could be obtained for all the crystalline samples and the amorphous samples of 2-d24 and 2-d90. To our satisfaction, spectra obtained for all crystalline samples of 2-d4, 2-d24, and 2-d90 indicated the presence of fast dynamic processes of the corresponding moieties that could be slowed (to a few kilohertz) when the samples were cooled (Figure 4). Those components were mostly static in amorphous samples, but they could be rendered mobile with sufficient thermal energy.

The sharp peak at the center of the spectrum of the crystalline samples of 2-d4 is the isotropic peak resulting from a small fraction of 2-d4 dissolved in residual solvent. The line shape of rest of the spectra could be simulated with a model that considers a 2-fold flip for the phenylene group. The best simulation at 293 K suggested a rotational frequency of 7.0 MHz. Rotational frequencies estimated in a similar manner for spectra measured at 283, 273, 263, 253, and 243 K indicated closest matches at 3.2 MHz, 1.6 MHz, 400 kHz, 180 kHz, and 25 kHz, respectively. The spectra obtained for the amorphous sample of 2-d4 turned out to be very noisy, even after over 48 h of acquisition time, due to the combined effects of the low deuterium content (only four deuterium atoms in one molecule) and the lower sensitivity of SS ^{2}H NMR in amorphous samples. Next we examined the 2H NMR spectra of 2-d24 at various temperatures. The spectra obtained for the crystalline sample could be simulated well by considering a 2 fold-flip model with a log-Gaussian distribution[18,21] of the rotational frequencies having a width $\sigma = 0.5$. Rates of rotation at 293, 283, 273, 263, and 253 K were estimated to be 3.1 MHz, 780 kHz, 300 kHz, 140 kHz, and 39 kHz, respectively. The need for a Gaussian distribution may be attributed to the coexistence of three crystallographically nonequivalent branch phenylenes in the sample. The ^{2}H NMR spectra

for amorphous sample of 2-d24 measured at 303 K showed a static powder pattern, indicating that motions in a glassy state are slower. As the amorphous sample was heated, line shape changes were also consistent with a 2-fold flipping process, just like the crystalline sample, but with a larger distribution of exchange rates ($\sigma = 2$). Rotational exchange frequencies for experimental spectra measured at 453, 403, 363, 333, and 303 K corresponded to 6.1 MHz, 520 kHz, 95 kHz, 11 kHz, and 2.0 kHz.

The ^{2}H NMR spectra of the peripheral phenyl groups measured with samples of 2-d90 are characterized by a rather complex superposition of a static signal, corresponding to the deuterium atom at the para-position, plus the signal of the other four deuterium atoms that are ortho and meta to the point of attachment of the phenyl ring. Since the C−D bond vector of the para-deuterium is aligned with the rotational axis (C−Ph bonds), its orientation does not change with rotation and its signal remains constant. By contrast, the other four C−D bonds have a cone angle of ±60° with respect to the rotational axis, such that any angular displacement can cause changes in the line shape of the spectrum. As shown in the right column in Figure 4, the spectral data of 2-d90 started with a spectrum approaching the slow exchange regime at 263 K and displayed spectral changes characteristic of increasing motion as the sample temperature reached 313 K. Notably, a simple 180° site exchange model could not provide a reasonable simulation for the experimental spectra. Instead, the experimental data could be reproduced reasonably well with a 4-fold rotation model and a 2:1 population distribution on nonequivalent sites related by 90°. This model led to site exchange rates of 2.0 MHz, 1.1 MHz, 690 kHz, 250 kHz, 100 kHz, and 30 kHz at 313, 303, 293, 283, 273, and 263 K, respectively.

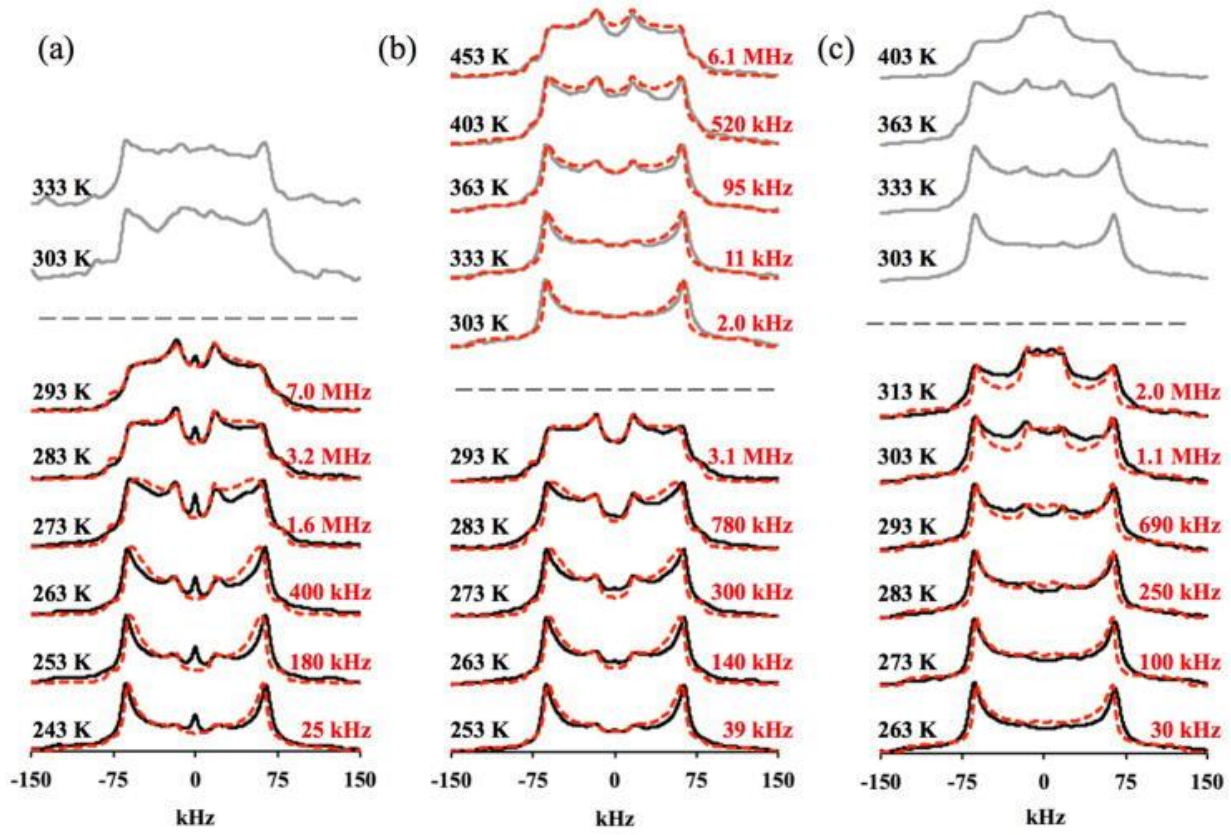


Figure 7.4. Experimental (black solid lines for the crystalline samples and grey solid lines for the amorphous samples) and simulated (red dashed lines) SS 2H NMR spectra of (a) rotor 2-d4, (b) rotor 2-d24, and (c) rotor 2-d90 at variable temperatures.

This model is an approximation because the real system has nine noncrystallographically equivalent peripheral phenyl groups, each with a potentially different site exchange rate. It is not surprising that the same 4-fold rotation model with a 2:1 population could only provide a qualitative simulation for the spectra of the amorphous 2-d90 samples with all phenyl groups rotating independently. This also suggests that rotational trajectories are dictated by molecular structures while rotational frequencies and their temperature dependence are determined by the crystallinity of the samples.

While it would be insightful to document the dynamics of the solvent, we were not able to observe them directly using NMR spectroscopy. Because the crystals are fragile and tend to lose solvent, all measurements had to be carried out in the presence of excess of solvent supernatant.

As a result, the signals corresponding to solvent molecules in the crystal lattice could not be identified.

7.4.4 Mechanisms of Rotation of Branch Phenylene and Peripheral Phenyl Groups: Correlated Trityl Group Rotations.

While the rotational dynamics of the central phenylene in 2-d4 are well accounted for by a 2-fold site exchange involving 180° rotations between degenerate sites, the branch phenylenes in 2-d24 and the peripheral phenyl groups in 2-d90 are part of trityl groups and could be expected to undergo correlated (gearing) motions. The fact that suitable simulations for 2-d24 required only a 2-fold site exchange model while those for 2-d90 required a 4-fold site exchange with unequal populations was investigated. We examined the rotational mechanisms for compound 2 with molecular dynamics (MD) simulations using the AMBER14 program.22 The parameters were generated with the antechamber module using the general Amber force field (GAFF) and atomic partial charges were assigned using the AM1-BCC method. Activation free energies in vacuum were obtained for different rotational mechanisms using umbrella sampling, and potentials of mean force (PMF) were constructed using the weighted histogram analysis method (WHAM), with the dihedral angles extracted every 50 fs.23 As suggested by the space-filling model (Figure 5a), a 2-fold rotation of one phenylene while the other two are completely static is an energetically demanding process because of the proximity of the three phenylene groups. With a calculated barrier of ca. 19 kcal/mol in the vacuum, it is unlikely to occur in the solid state. However, it is possible to have independent rotation of one ring while the other two are oscillating, and the corresponding activation energy was estimated to be only 3−5 kcal/mol in vacuum. Alternatively, the three rings could rotate synchronously to change the absolute configuration of the chiral propeller conformation of the trityl group, a process extensively studied by Mislow and others.24

In vacuum, this enantiomerization process is associated with angular displacements of ca. 90° and energy barriers of about 1 kcal/mol (Figure 5b, dashed curves). Analysis of the lowest energy trajectory suggests a "2-ring flip" mechanism, as proposed by Mislow for trityl isomerization in solution. In the crystal, however, the two enantiomers (P and M) are in a chiral environment and have different energies. Consequently, if one enantiomer is much higher in energy in the crystal (Figure 5b, blue curve), an effective 2-fold rotation (from one P-enantiomer to a transient M enantiomer to the next P-enantiomer) would be observed, instead of a 4-fold rotation. Under the most extreme circumstances, one enantiomer could be a local maximum (transition state of rotation) and the correlated motion of three phenylenes would also appear to be 2-fold rotations (Figure 5b, green curve). As a result, while the experimental data and MD simulation are both consistent with an effective 2-fold rotation in the case of 2-d24, it is not possible to distinguish between correlated motion of three phenylenes and the independent rotation of one phenylene while the other two oscillate.

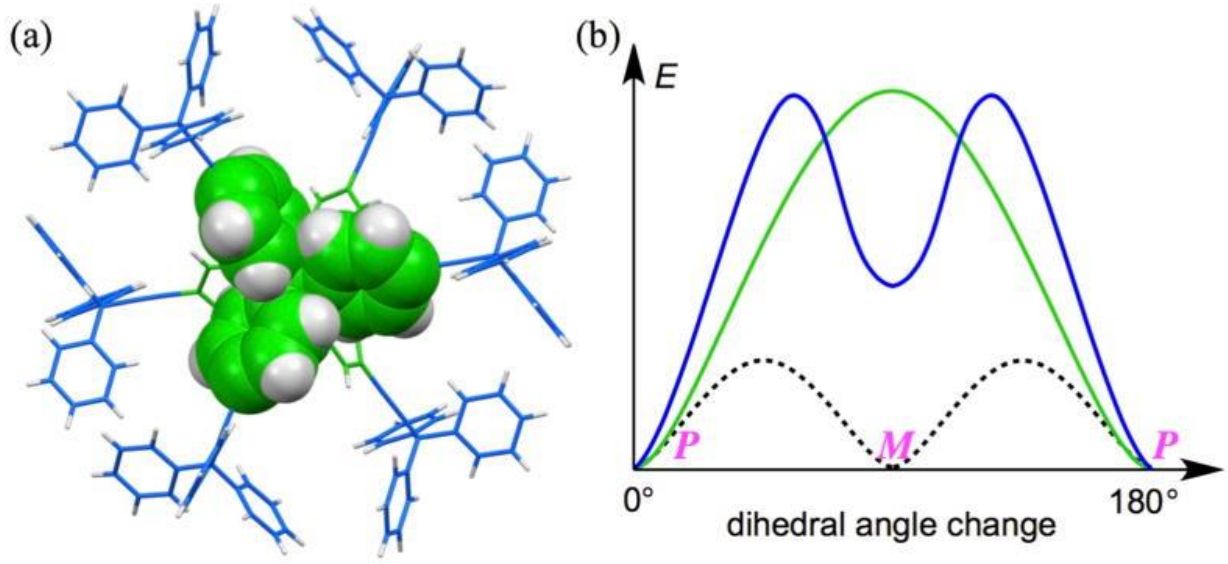


Figure 7.5. (a) Space filling representation of a branch phenylene trityl group indicating the necessity of a correlated process to allow for two-fold 180° rotations. (b) Energy changes of correlated phenylene rotation in a trityl group in vacuum corresponding for apparent two fold (green solid curve) and four fold (blue solid curve and black dashed curve) rotation modes.

The mechanisms of rotation for the peripheral phenyl groups in 2-d90 samples can be analyzed also as discussed above and considering the potentials shown in Figure 5b. In fact, the experimentally observed 4-fold rotational model can be understood in terms of a trityl enantiomerization process where the crystallographically observed enantiomer is lower in energy but the second enantiomer is thermally accessible. Thus, a ca. 2:1 population distribution suggested that the two enantiomers (P and M) are both local minima and the energy difference between them is small (Figure 5b, blue curve). This is also consistent with the crystal structure, where the periphery phenyls displayed larger thermal ellipsoids than the rest of the rotor structure.

7.4.5 Activation Parameters.

The Arrhenius plots constructed from the rotational exchange frequencies observed at each of the experimental temperatures for the central phenylene (2-d4), branch phenylenes (2-d24), and peripheral phenyls (2-d90) are shown in Figure 6. The most significant result is that all the energies (Ea) and pre-exponential factors (A) are 15.7 kcal/mol and $3.5 \times 1018\ s^{-1}$ for 2-d4, 15.5 kcal/mol and $7.8 \times 1017\ s^{-1}$ for 2-d24, and 13.7 kcal/mol and $7.8 \times 1015\ s^{-1}$ for 2-d90. Similar activation parameters obtained for different aromatic segments in the molecule are consistent with the fact that all of them share contacts and thus have a shared potential (Figure 3).

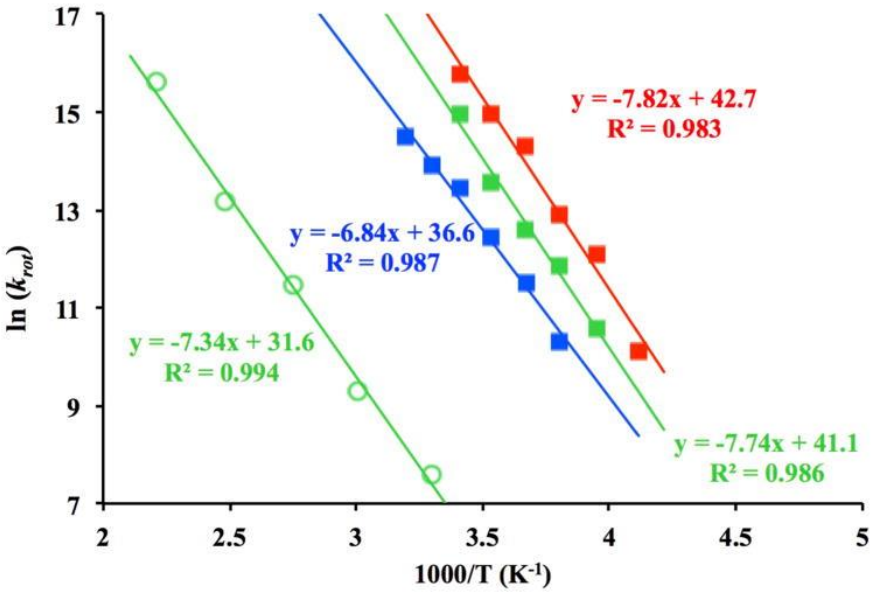


Figure 7.7. Arrhenius plots for crystalline (filled squares) and amorphous samples (open circles) of isotopologues of 2.

It is important to note that pre-exponential factors on the order of 1016−1018 s−1 cannot be associated with an elementary process, such as the torsional mode that becomes an internal rotation when provided with sufficient thermal energy. In fact, elementary rotations involving phenylene groups are limited to values on the order of ca. 1012 s^{-1}, as shown by quantum mechanical calculations of the corresponding torsional mode or by classical mechanics calculations based on its moment of inertia.25 Numerous previous studies with crystalline phenylene rotors have also revealed pre-exponential values of this magnitude, with exceptions encountered when noncrystalline samples are analyzed.26 We have previously suggested that abnormally high pre-exponential factors in the solid state can be associated with changes in the fluidity of the structure, rather than with a static potential of the corresponding motions. For example, when exploring the temperature dependence of the rotational dynamics of a p-phenylene group in a periodic mesoporous organosilicate (PMO), we obtained nonlinear data within a temperature region indicating an apparent activation energy of 47 kcal/mol and an apparent pre-exponential value of 4.0 × 1041 s^{-1}, both of which are, of course, nonsensical.26 Further analysis

based on differential scanning calorimetry (DSC) supported a reasonable interpretation for the observed results. It was shown that the steep slope of the rotational frequency vs inverse temperature in the corresponding Arrhenius plot was the result of the structural softening that occurs when the 2D rigid glass becomes a 2D rotational fluid, during a second order glass transition. In a similar manner, we propose that a large slope and intercept in the Arrhenius plot of crystalline dendrimeric molecular rotor 2 is the result of a crystal structure that becomes increasingly soft, or more "fluidlike", as the temperature increases. We suggest the term crystal fluidity to convey the softening of the local environment in crystals that results from the temperature-dependent conformational motions, librations, group rotations, and presumably solvent dynamics.

If we assume that the correct pre-exponential factor for any given potential is dictated by the torsional motion of a phenylene with a constant value of ca. $1.0 \times 1012\ s^{-1}$, we can estimate a "corrected" activation energy for every temperature. In the case of 2-d4, the experimental data would be consistent with an energy barrier that changes from 7.9 kcal/mol at 253 K, when the structure is relatively rigid, to a barrier of 6.9 kcal/mol at 293 K as the structure becomes more fluid. A similar analysis for the data obtained in the case of 2-d24 gives what would be "corrected" activation energies of 8.6 and 7.4 kcal/mol at 253 and 293 K, respectively. Thus, it is reasonable to conclude that the dynamic behavior observed with crystals of 2-d4 and 2-d24 is the result of a barrier change of ca. 1 kcal/mol over a rather modest temperature range of only 40 K. A similar analysis with samples of amorphous 2-d24 showed that variations in the "corrected" activation energy over a similar temperature interval are much smaller. Furthermore, an interpretation based on a decreasing barrier is also consistent with the fact that all aromatic groups are undergoing

temperature-activated, fast rotational dynamics, such that their local potential becomes increasingly fluid.

Table 7.1. Summary of apparent[#] activation parameters. [#]Please see discussion in main text.

Sample	E_a **(kcal/mol)**	A **(s^{-1})**
2-d_4, crystalline	15.7	$3.5x10^{18}$
2-d_{24}, crystalline	15.5	$7.8x10^{17}$
2-d_{24}, amorphous	14.7	$5.3x10^{13}$
2-d_{90}, crystalline	13.7	$7.8x10^{15}$

7.5 Conclusion

A highly efficient convergent strategy was developed for the synthesis of a macromolecular dendritic rotor in its natural abundance and selectively deuterated forms to carry out quadrupolar echo ^{2}H NMR line shape analysis. Amphidynamic crystals obtained from a mixture of 2,4,6-trimethylpyridine and 2,2,4-trimethylpentane showed a low-density structure that allows for the fast rotation of all the aromatic groups, including one located at the core, six at the branches, and 18 at the periphery. It is notable that the rotational frequencies of the three segments of the structures reach the megahertz regime near ambient temperatures and exhibit similar activation parameters. While close activation energies are consistent with rotational potentials based on interactions between the three parts of the structure, pre-exponential factors that are up to 6 orders of magnitude larger than those expected for elementary processes are interpreted in terms of a temperatureinduced softening of the packing structure. These observations indicate that rotational dynamics in this hyperbranched structure depend more on the fluidity of the crystal than on the contacts and interactions that would make up a static potential derived from the equilibrium position of the atoms in the structure.

7.6 References

1. (a) Abendroth JM, Bushuyev OS, Weiss PS, Barrett CJ. ACS Nano. 2015;9:7746. (b) Erbas-Cakmak S, Leigh DA, McTernan CT, Nussbaumer AL. Chem Rev. 2015;115:10081. (c) Coskun A, Banaszak M, Astumian RD, Stoddart JF, Grzybowski BA. Chem Soc Rev. 2011;41:19. (d) Kinbara K, Aida T. Chem Rev. 2005;105:1377. (e) Kottas GS, Clarke LI, Horinek D, Michl J. Chem Rev. 2005;105:1281.

2. (a) Cheng C, McGonigal PR, Schneebeli ST, Li H, Vermeulen NA, Ke C, Stoddart JF. Nat Nanotechnol. 2015;10:547. (b) Li Q, Fuks G, Moulin E, Maaloum M, Rawiso M, Kulic I, Foy JT, Giuseppone N. Nat Nanotechnol. 2015;10:161. (c) Ragazzon G, Baroncini M, Silvi S, Venturi M, Credi A. Nat Nanotechnol. 2015;10:70. (d) De S, Pramanik S, Schmittel M. Angew Chem Int Ed. 2014;53:14255. (e) Conyard J, Cnossen A, Browne WR, Feringa BL, Meech SR. J Am Chem Soc. 2014;136:9692. (f) Lewandowski B, De Bo G, Ward JW, Papmeyer M, Kuschel S, Aldegunde MJ, Gramlich PME, Heckmann D, Goldup SM, D'Souza DM, Fernandes AE, Leigh DA. Science. 2013;339:189. (g) Perera UGE, Ample F, Kersell H, Zhang Y, Vives G, Echeverria J, Grisolia M, Rapenne G, Joachim C, Hla SW. Nat Nanotechnol. 2013;8:51.(h) Thies S, Sell H, Schütt C, Bornholdt C, Näther C, Tuczek F, Herges R. J Am Chem Soc. 2011;133:16243.

3. (a) Vogelsberg CS, Garcia-Garibay MA. Chem Soc Rev. 2012;41:1892. (b) Khuong TAV, Nuñez JE, Godinez CE, Garcia-Garibay MA. Acc Chem Res. 2006;39:413. (c) Garcia-Garibay MA. Proc Natl Acad Sci USA. 2005;102:10771.

4. (a) Pérez-Estrada S, Rodríguez-Molina B, Xiao L, Santillan R, Jiménez-Osés G, Houk KN, Garcia-Garibay MA. J Am Chem Soc. 2015;137:2175. (b) Jiang X, Rodríguez-Molina B, Nazarian N, Garcia-Garibay MA. J Am Chem Soc. 2014;136:8871. (c) Commins P, Garcia-Garibay MA. J Org Chem. 2014;79:1611. (d) Rodríguez-Molina B, Farfán N, Romero M, Mendez-Stivalet MJ,

Santillan R, Garcia-Garibay MA. J Am Chem Soc. 2011;133:7280. (e) Jarowski PD, Houk KN, Garcia-Garibay MA. J Am Chem Soc. 2007;129:3110.

5. (a) O'Brien ZJ, Natarajan A, Khan S, Garcia-Garibay MA. Cryst Growth Des. 2011;11:2654.(b) Karlen SD, Reyes H, Taylor RE, Khan SI, Hawthorne MF, Garcia-Garibay MA. Proc Natl Acad Sci U S A. 2010;107:14973. (c) Rodríguez-Molina B, Ochoa ME, Farfán N, Santillan R, Garcia-Garibay MA. J Org Chem. 2009;74:8554. (d) Khuong TAV, Dang H, Jarowski PD, Maverick EF, Garcia-Garibay MA. J Am Chem Soc. 2007;129:839. (e) Dominguez Z, Khuong T, Dang H, Sanrame CN, Nuñez JE, Garcia-Garibay MA. J Am Chem Soc. 2003;125:8827. (f) Dominguez Z, Dang H, Strouse MJ, Garcia-Garibay MA. J Am Chem Soc. 2002;124:7719. (g) Dominguez Z, Dang H, Strouse MJ, Garcia-Garibay MA. J Am Chem Soc. 2002;124:2398.

6. (a) Harada J, Ohtani M, Takahashi Y, Inabe T. J Am Chem Soc. 2015;137:4477. (b) Zhu K, O'Keefe CA, Vukotic VN, Schurko RW, Loeb SJ. Nat Chem. 2015;7:514. (c) Nawara Hultzsch AJ, Stollenz M, Barbasiewicz M, Szafert S, Lis T, Hampel F, Bhuvanesh N, Gladysz JA. Chem Eur J. 2014;20:4617. (d) Comotti A, Bracco S, Yamamoto A, Beretta M, Hirukawa T, Tohnai N, Miyata M, Sozzani P. J Am Chem Soc. 2014;136:618. (e) Zhang QC, Wu FT, Hao HM, Xu H, Zhao HX, Long LS, Huang RB, Zheng LS. Angew Chem Int Ed. 2013;52:12602. (f) Lemouchi C, Iliopoulos K, Zorina L, Simonov S, Wzietek P, Cauchy T, Rodríguez-Fortea A, Canadell E, Kaleta J, Michl J, Gindre D, Chrysos M, Batail P. J Am Chem Soc. 2013;135:9366. (g) Zhang W, Ye HY, Graf R, Spiess HW, Yao YF, Zhu RQ, Xiong RG. J Am Chem Soc. 2013;135:5230. (h) Vukotic VN, Harris KJ, Zhu K, Schurko RW, Loeb SJ. Nat Chem. 2012;4:456. (i) Akutagawa T, Koshinaka H, Sato D, Takeda S, Noro SI, Takahashi H, Kumai R, Tokura Y, Nakamura T. Nat Mater. 2009;8:342.

7. (a) Schliwa M, Woehlke G. Nature. 2003;422:759. (b) Piccolino M. Nat Rev Mol Cell Biol. 2000;1:149.

8. (a) May R, Jester SS, Höger S. J Am Chem Soc. 2014;136:16732. (b) Shen X, Ho DM, Pascal RA. J Am Chem Soc. 2004;126:5798.

9. Karlen SD, Garcia-Garibay MA. Chem Commun. 2005:189.

10. Heck H, Simon RL, Anbar M. J Chromatogr. 1977;133:281.

11. Tanaka N, Thornton ER. J Am Chem Soc. 1976;98:1617.

12. (a) Kondratuk DV, Sprafke JK, O'Sullivan MC, Perdigao LMA, Saywell A, Malfois M, O'Shea JN, Beton PH, Thompson AL, Anderson HL. Chem Eur J. 2014;20:12826. (b) Bauer RE, Enkelmann V, Wiesler UM, Berresheim AJ, Müllen K. Chem Eur J. 2002;8:3858. (c) Nielsen MB, Schreiber M, Baek YG, Seiler P, Lecomte S, Boudon C, Tykwinski RR, Gisselbrecht JP, Gramlich V, Skinner PJ, Bosshard C, Gunter P, Gross M, Diederich F. Chem Eur J. 2001;7:3263. (d) Ipaktschi J, Hosseinzadeh R, Schlaf P. Helv Chim Acta. 1998;81:1821.

13. Crystals of almost identical cell parameters could be obtained from an o-xylene/2,2,5-trimethylpentane mixture with the only differences being the solvent molecules included in the crystals.

14. Spek A. Acta Crystallgr D. 2009;65:148.

15. The packing coefficient increases to 0.69 when the solvent molecules in the crystal are included. Molecular volumes were calculated using the group increment approach reported in Gavezzotti A. J Am Chem Soc. 1983;105:5220.Bondi A. J Phys Chem. 1964;68:441.. The results agree closely with volumes calculated using (c) Spartan'08 Wavefunction, Inc. Irvine, CA.

16. Dunitz JD, Gavezotti A. Acc Chem Res. 1999;32:677.

17. water in proteins

18. Hansen MR, Graf R, Spiess HW. Acc Chem Res. 2013;46:1996.

19. Macho V, Brombacher L, Spiess HW. Appl Magn Reson. 2001;20:405.

20. The SS 2H NMR spectra for the crystalline samples were obtained in the presence of the solvent supernatant to avoid solvent loss and amorphization during the data acquisition. For all the VT experiments, samples were gradually heated up or cooled down from 293 K to the target temperature. Sample integrity and data reproducibility was analyzed by measureing the 293 K spectra at the end of each experiment.

21. Case DA, Babin V, Berryman J, Betz RM, Cai Q, Cerutti DS, Cheatham TE, III, Darden TA, Duke RE, Gohlke H, Goetz AW, Gusarov S, Homeyer N, Janowski P, Kaus J, Kolossvary I, Kovalenko A, Lee TS, LeGrand S, Luchko T, Luo R, Madej B, Merz KM, Paesani F, Roe DR, Roitberg A, Sagui C, Salomon-Ferrer R, Seabra G, Simmerling CL, Smith W, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Kollman PA. AMBER 14. 2014

22. Grossfield A. [accessed Jan 27, 2016];Wham: the weighted histogram, analysis method, version 2.0.9. http://membrane.urmc.rochester.edu/content/wham.

23. Gust D, Mislow K. J Am Chem Soc. 1973;95:1535.

24. Inertial rotational frequency τIR−1 can be estimated from the moment of inertia (I) of a rotator using the equation τIR−1 = [(2π/9) (I/kBT)1/2]−1, where I is moment of inertia measured along the rotational axis. Please see: Kawski A. Crit Rev Anal Chem. 1993;23:459.

25. Vogelsberg CS, Bracco S, Beretta M, Comotti C, Sozzani P, Garcia-Garibay MA. J Phys Chem B. 2012;116:1623.

Chapter 8. Ultrafast Rotation in an Amphidynamic Crystalline Metal Organic Framework

8.1 Abstract

Amphidynamic crystals are an emergent class of condensed phase matter designed with a combination of lattice-forming elements linked to components that display engineered dynamics in the solid state. Here, we address the design of a crystalline array of molecular rotors with inertial diffusional rotation at the nano scale, characterized by the absence of steric or electronic barriers. We solved this challenge with 1,4-bicyclo[2.2.2]octane dicarboxylic acid (BODCA)MOF, a metal-organic framework (MOF) built with a high-symmetry bicyclo[2.2.2]octane dicarboxylate linker in a Zn_4O cubic lattice. Using spin-lattice relaxation ^1H solid-state NMR at 29.49 and 13.87 MHz in the temperature range of 2.3–80 K, we showed that internal rotation occurs in a potential with energy barriers of 0.185 kcal·mol^{-1}. These results were confirmed with ^2H solid-state NMR line-shape analysis and spin-lattice relaxation at 76.78 MHz obtained between 6 and 298 K, which, combined with molecular dynamics simulations, indicate that inertial diffusional rotation is characterized by a broad range of angular displacements with no residence time at any given site. The ambient temperature rotation of the bicyclo[2.2.2]octane (BCO) group in BODCA-MOF constitutes an example where engineered rotational dynamics in the solid state are as fast as they would be in a high-density gas or in a low-density liquid phase.

8.2 Introduction

Crystalline arrays of barrierless molecular rotors present unique opportunities to explore emergent phenomena while offering opportunities for the development of novel paradigms in materials science and engineering [(1)]. While the juxtaposition of high kinetic energy and long-range molecular order is uncommon, order in motion provides the required anisotropy for the control of thermal, optical, and dielectric properties, as well as for the design of dipolar arrays [(2)]

and the development of artificial molecular machines [3]. While some progress has been made in the study of experimental 2D rotary systems by taking advantage of molecules bound to surfaces [4, 5], barrierless molecular rotors in 3D crystals have not been available up to now. The first step to meet this challenge is to identify molecular components with intrinsic (gas phase) rotational barriers that are smaller than thermal energy (i.e., Ea $\leq$0.596kcal·mol^{-1} at 300 K), and crystal structures that poseno hindrance to the rotating units.The four key architectural elements to consider are (i) the size and shape of the rotator, which determine its surface area and moment of inertia, (ii) the availability of free volume with a cross-section that is greater than the volume of revolution of the rotator, (iii) the chemical nature and intrinsic torsional barrier of the bond or axle that links the rotator to the stator, and (iv) the axial symmetries of the rotator and the stator. Based on recent observations [6–8], we selected bicyclo[2.2.2] octane (BCO) as a promising rotator for the construction of a crystalline molecular rotor with a barrier that is smaller than thermal energy at 300 K (Fig. 1, Right). In agreement with its relatively small size, globular shape, and small moment of inertia, molecular [7, 8] and supramolecular [9] crystals built with BCO rotators have very low activation energies (Ea~1.15–2.7kcal·mol^{-1}), and their preexponential factors (attempt frequencies) approach the frequency of the torsional mode that turns into rotations (~1012 s^{-1}) [6–8]. For the selection of a crystal lattice that can generate an empty volume that is greater than the volume of revolution of the BCO rotator (~5.6-Å diameter), we decided to explore the use of metal-organic frameworks (MOFs) [9, 10] because of their ability to maintain relatively large and permanent pores in highly robust architectures [11, 12].

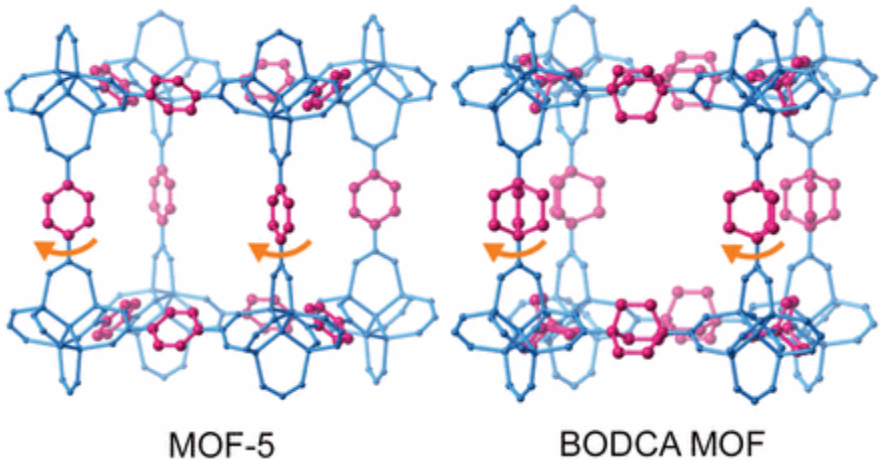


Figure 8.1. Isostructural networks of MOF-5 and BODCA-MOF with the corresponding BDC and BODCA rotators shown in red and the static carboxylate and Zn4O clusters shown in blue (hydrogen atoms omitted for clarity).

8.3 Materials and Methods

Rotator Synthesis.

All of the chemicals and solvents used in this work and their sources and purities are described in detail in SI Appendix. Also included are the full synthetic procedure for the bicyclo[2.2.2]octane dicarboxylic acid in its natural abundance, and deuterated forms as well as their detailed spectroscopic and mass spectrometric characterization are described in detail in SI Appendix.

BODCA-MOF Synthesis and Characterization.

Microcrystalline samples of BODCA-MOF were obtained in DMF at room temperature by mixing solutions of BODCA and zinc acetate dehydrate and letting the mixture stand for 15 h. Crystals were separated by centrifugation and DMF was exchanged by CH_2Cl_2, which was subsequently removed under reduced pressure over night at room temperature followed by heating at 40 °C for 2 h. Solid samples were characterized by PXRD, solid-state ^{13}C CPMAS NMR, thermal analysis, and gas sorption measurements. White microcrystalline powders prepared for ^{1}H T_1 and ^{2}H spin-echo wide-line solid-state NMR experiments were back-filled with He(g).

Solid-State NMR Spectroscopy and Dynamics.

^{1}H NMR spin-lattice relaxation (T_1) measurements in the solid state were carried out at University of California, Los Angeles at two different ^{1}H Larmor frequencies of 29.49 and 13.87 MHz and over a wide range of temperatures. Wide-line ^{1}H spectra from 2.3 to 80 K were measured in a microcrystalline BODCA-MOF sample sealed in a capillary under an ultrahigh-purity helium atmosphere. ^{2}H solid-state NMR wide-line spectra and Zeeman spin-lattice relaxation times were acquired at Pacific Northwest National Laboratory's (PNNL) Environmental and Molecular Sciences High Field Magnetic Resonance Facility at a Larmor frequency of 76.78 MHz in the temperature range of 6 to 298 K.

8.4 Results and Discussion

Low-Barrier, High-Symmetry Rotator

Recognizing the similar dimensions of BCO and benzene (Fig. 2A), it seemed reasonable that one should be able to prepare an MOF-like structure based on octahedral Zn4O nodes. Instead of using1,4-benzenedicarboxylate (BDC) as the network linker giving rise to the well-known MOF-5, crystals could be built with 1,4-bicyclo[2.2.2]octane dicarboxylic acid (BODCA). While it is known that the rotation of the 1,4-phenylene in MOF-5 has a relatively large energy barrier of ∼11.3 ± 2 kcal mol −1 as a result of loss of conjugation between the benzene ring and the two carboxylates, which renders the transition state a second-order saddle point [(9)], the absence of π-conjugation between the saturated BCO framework and the carboxylate group generates a low barrier to rotation in the case of BODCA-MOF. Furthermore, the threefold axial symmetry order of the BCO rotator (CR = 3) combines with the twofold symmetry order of the static carboxylate (CS = 2) to generate a six-fold degenerate potential (Fig.2B). Quantum-mechanical calculations indicate that energy minima have an eclipsing interaction between the plane of the carboxylate

group and one of the BCO blades, such that degenerate structures are obtained by rotation of the corresponding dihedral every 60°. Energy scans carried out with different methods and basis sets indicated very low-energy barriers. For example, calculations using the hybrid functionals M062X/6–311+G(d) or ωB97XD/6–311+G(d) indicated that the energy barrier is only 0.1 kcal mol−1. By comparison, rotational barriers calculated with MP2/6–31G(d) or B3LYP/6–31G(d) are 0.4 and 0.5 kcal·mol^{-1}, respectively. Based on the large empty space of the MOF framework, the relatively small angular steps between adjacent minima (60°), and the cylindrical shape of the BCO rotator, it seems reasonable to expect that rotational barriers should be similar in the solid state.

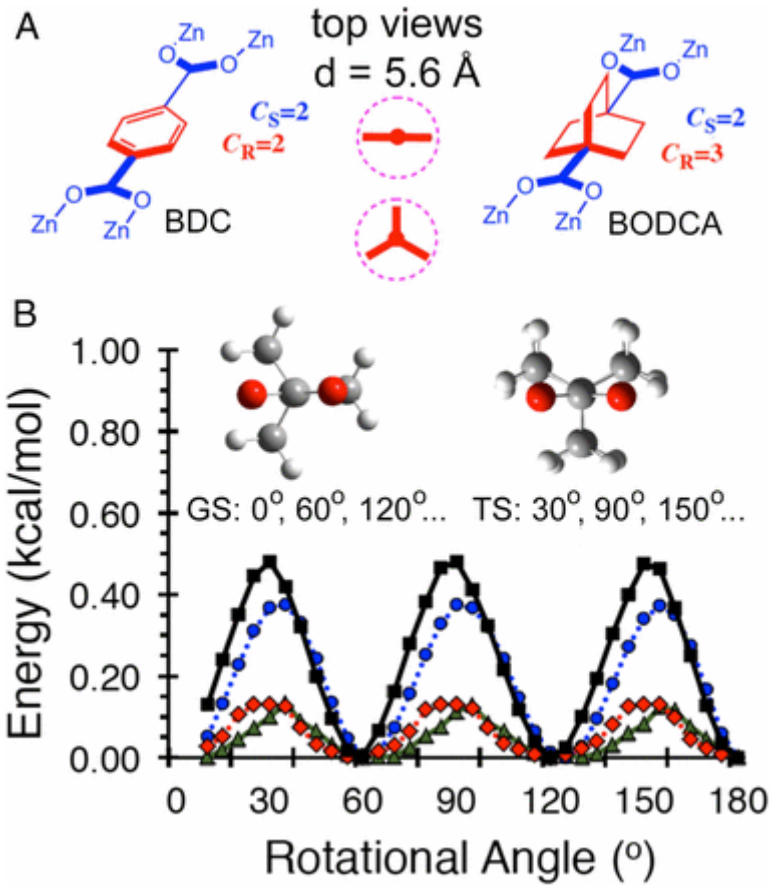


Figure 8.2. (A) Line structures of BDC, BODCA, and a representation of the cross-section of their volumes of revolution viewed down the main molecular axes. The diameters of the two rotors (d) are 5.6 Å. (B) Results from energy scans illustrating half a period (180°) of the sixfold degenerate energy potential of BODCA-MOF with the ground-state (GS) and transition-state (TS) structures

viewed down the BCO-carboxylate linkage [black squares B3LYP/6–31G(d); blue circles MP2/6–31G(d); red rhombs ωB97xd/6–311+G(d); green triangles, M062X/6–311+G(d)]. The rotational angle is defined by the dihedral formed between the pane of the carboxylate and the plane of the BCO blade. GS geometries have dihedral angles of 0, 60, 120, etc.; the TS have dihedral angles of 30, 90, 150, etc. High-temperature measurements with RT > Ea are expected to reflect diffusional rotation, and low-temperature measurements with RT < Ea are expected to occur by a site-exchange mechanism.

Synthesis and Characterization of BODCA-MOF and BODCA-d-MOF

To test our hypothesis, we prepared samples of BODCA as reported in the literature [(11)]. The synthesis of deuteronenriched BODCA-d ligand for dynamic measurements using quadrupolar echo ^{2}H NMR measurements was accomplished as shown in Fig. 3 by taking advantage of a base-catalyzed H/D exchange during the Huang–Minlon modified Wolff–Kishner reduction [(13)] of dione 3 by the introduction of triethylene glycol-d2 and NaOD. This procedure resulted in a statistical Dlabeling distribution established through a combination of liquid field desorption ionization mass spectrometry and the integration of the ^{1}H solution NMR spectrum of deuterated diol 4. On average, three out of the eight exchangeable protons were replaced with deuterons, resulting in a typical D content of 3 out of the 12 total hydrogen positions (four positions are not exchangeable), or 25%, on the bicyclic rotator. As expected, the ^{1}H and ^{13}C solution NMR spectra of diol 4-d and diacid BODCAd display chemical shifts and coupling with isotope effects due to the various substitution patterns of the deuterons with respect to the NMR-active ^{1}H and ^{13}C nuclei.

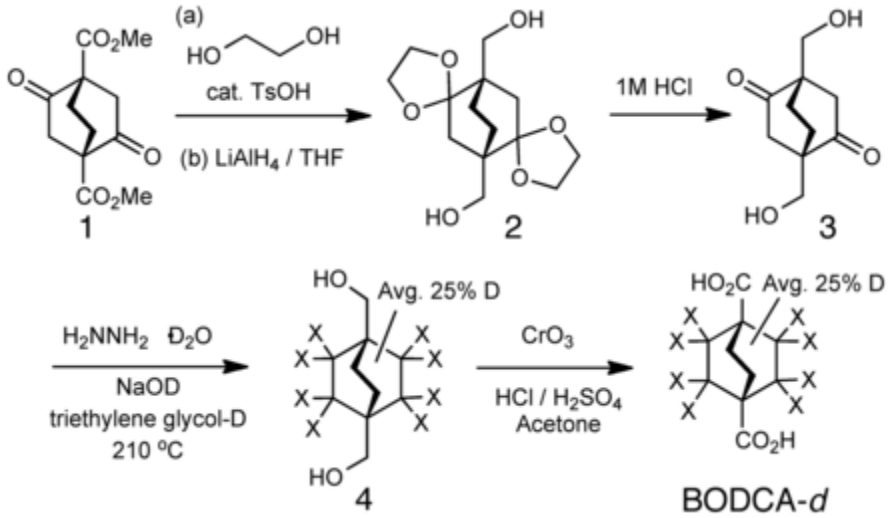


Figure 8.3. Synthesis of the BODCA ligand with modifications to install deuterons on the bicyclic core, X represents H or D atoms.

Crystals of BODCA-MOF and BODCA-d-MOF were obtained via a room-temperature synthesis. Separate solutions of zinc acetate dehydrate and the BODCA ligand (either natural abundance for ^{1}H NMR studies, or deuteron enriched for ^{2}H NMR studies) in N, N-dimethyl formamide (DMF) were mixed together in a vigorously stirring solution. Upon initial mixing, a white precipitate was immediately observed, suggesting the formation of zinc oxidecarboxylate clusters leading to MOF assembly. Aging of the sample in the mother liquor followed by successive washings (DMF and chloroform) and removal of solvent by heating under vacuum to create empty cores resulted in the desired compoundas a white microcrystalline powder. Thermogravimetric analysis of the heat-treated solid confirmed the lack of included solvent and thestructural integrity ofthedesolvated sample. The dinitrogen adsorption isotherm (77 K) of BODCA-MOF displays a sharp uptake with a single adsorption step at p/po > 0.01 followed by saturation, indicating a microporous material. It can also be described as a type II isotherm under the International Union of Pure and Applied Chemistry (IUPAC) classification system. Application of the Brunauer– Emmett–Teller (BET) model over 0.005 > p/po > 0.05, according to the recommendation by Walton and Snurr [(14)], resulted in a BET surface area of SBET = 2,150 m2

g^{-1}. The total pore volume was determined from the Dubinin–Raduschevich (DR) model over the $0.13 > p/po > 0.35$ range, obtaining a $V_{p,DR} = 0.8630\ cm^3 \cdot g^{-1}$.

Analysis by ^{13}C cross-polarization and magic angle spinning (CP/MAS) solid-state NMR suggested a very high-symmetry structure, as indicated by the presence of only three narrow signals at 188.3, 40.4, and 28.2 ppm, which are assigned to the carboxylate, bridgehead, and methylene carbons of the BODCA linkers. Whereas the described synthesis protocol was unable to produce large crystals suitable for singlecrystal X-ray diffraction, the microcrystalline powder was studied using powder X-ray diffraction (PXRD) crystallography. BODCA-MOF displays a PXRD pattern that exhibits sharp Bragg peaks between 7° and 80° (2-theta, CuKα radiation) evidencing a high degree of crystallinity with a resolution up to 0.80 Å. Indexing of the diffraction pattern via the TREOR algorithm resulted in a cubic crystal system with cell parameter a TREOR = 12.678(29) Å [figures of merit M(13) = 17, F13 = 11 (0.021722, 59)], a value very close to MOF-5 in its primitive setting [15]. Extraction of the observed integrated intensities through the Pawley method using the cell parameter ($a' = 2$ a TREOR) and space groups P23 and P1, for a cubic supercell similar to MOF-5, allowed the calculation of low-resolution electron density reconstructions. Electron density maps were obtained using the chargeflipping algorithm integrated in the program SUPERFLIP. The reconstructed maps display regions of electron density that resemble the zinc-oxide clusters of MOF-5, suggesting the isoreticular nature of BODCA-MOF (sharing the same topology as MOF-5). Crystal modeling using the supercell parameter in Materials Studio Modeling Suite (Biovia v8.0) allowed the generation of a plausible crystal model (space group Pm-3) for the crystal solution using the Rietveld method in the general structure analysis system (GSAS) software [16]. Rietveld refinement of the experimental diffraction pattern using the modeled crystal allowed the solution of the MOF in the Fm-3m space group (to allow

for the high degree of positional disorder). The refined structure exhibits a considerable amount of positional disorder at the organic BCO moiety, suggesting a high dynamicbehavior of theaverage structure, in agreementwith the proposed cubic space group containing an axially substituted rotator undergoing fast rotation. This result is analogous to that observed with plastic crystals formed by globular molecules where an apparent increase in symmetry is observed as a result of whole-body rotation.

Temperature-Dependent Dynamic Mechanisms for SolidState 1H NMR Spin-Lattice Relaxation

As in previous studies [7–9], the rotational dynamics of the BCO rotator in the BODCA-MOF lattice were analyzed by temperature dependent solid-state ^{1}H NMR spin-lattice relaxation (T_1). The method relies on the fact that nuclear spin relaxation occurs via stimulated transitions arising from the modulation of dipolar magnetic interactions near the Larmor frequency ($\omega_0 = 2\pi\nu_0$) of the observed nucleus, which are caused by dynamic processes with correlation times (τ_c) assumed to follow a temperature-dependent Arrhenius behavior with an activation energy Ea and an attempt frequency (or preexponential factor) τ_0^{-1} (Eq. 1) [17]. The relation between the observed macroscopic spin-lattice relaxation T_1^{-1} and the molecular correlation times τc is given by the Kubo–Tomita equation (Eq. 2) [18, 19], which indicates that relaxation becomes more efficient when the condition $\tau_c^{-1} = \omega_0$ is met by the system. The definition of the constant C is given in Eq. 3. It represents the strength of the dipolar interactions involved in the relaxation process, whichisa function of the number of nuclei responsible for the relaxation (n) to the total number of nuclei that need to be relaxed (N), where μ_0 is the permeability of free space, γ is the gyromagnetic ratio, Z is the reduced Plank constant, and r is the internuclei distance.

$$\tau_c^{-1} = \tau_0^{-1} \exp\left(E_a/\mathrm{RT}\right),$$

([1])

$$\mathrm{T}_1^{-1} = \mathrm{C}\left[\tau_c\left(1+\omega_0^2\tau_c^2\right)^{-1} + 4\tau_c\left(1+4\omega_0^2\tau_c^2\right)^{-1}\right],$$

([2])

$$\mathrm{C} = (n/N)\,(9/40)\,(\mu_0/4\pi)^2\gamma^4\hbar^2/r^6.$$

([3])

In the case of BODCA-MOF, all ^{1}H atoms are able to contribute tothe ^{1}H T^1 relaxation process. The distance between 1Hatomsin adjacent rotators is, however, relatively large ($r \geq 4$ Å) and gives rise to weak dipolar fields ($B_{nuc} = [\gamma h \mu_0/(4\pi r^3)] \leq 0.44$ Gauss), such that relaxation should occur primarily by intrarotor interactions ($r \geq 2.2$ Å) giving rise to much greater dipolar fields ($B_{nuc} \sim 2.5$ G). Consequently, the two most important intramolecular degrees of freedom are a low-energy conformational twisting process and the rotation of interest about the BCO 1,4-axis (Fig. 4A). As shown in Fig. 2B, the rotational energy profile about the BCO dicarboxylate linkage is very shallow with activation energies on the order of ~0.1–0.5 kcal·mol^{-1}, such that that rotation can contribute to the nuclear spin relaxation at sufficiently low temperatures, when it slows down into the megahertz regime.

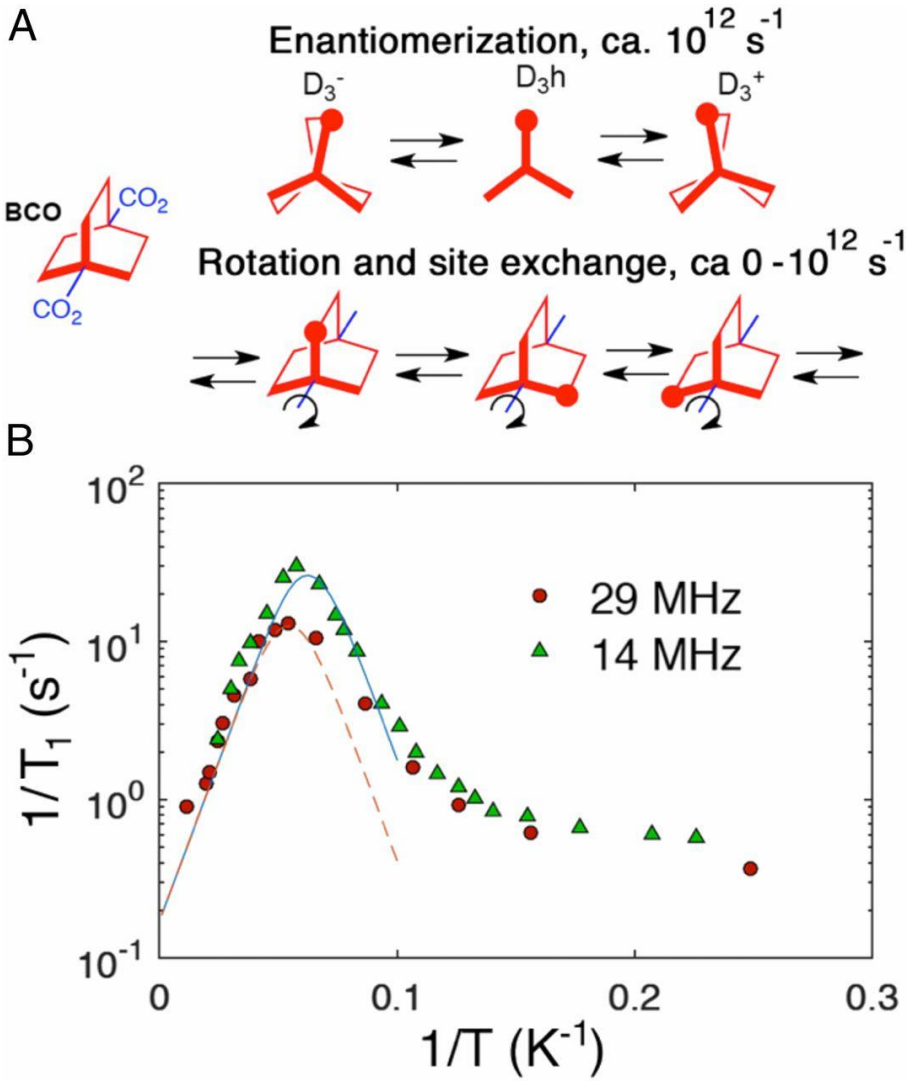


Figure 8.4. (A) Dynamic processes of the 1,4-BCO rotators that may contribute to the spin-lattice relaxation of the 1H nuclei. Conformational twisting has the characteristics of a normal mode with frequencies that are too high (~1012 s^{-1}) to contribute to ^1H spin-lattice relaxation in the megahertz (106–108 s^{-1}) regime. By contrast, inertial rotation and rotation by a site-exchange mechanism has a small barrier and may be slowed down to megahertz regime and constitute an efficient spin-lattice relaxation mechanism. (B) ^1H T_1 relaxation data of the natural abundance BODCA-MOF from T = 2.3–80 K, plotted with a semilog scale for T_1^{-1} vs. 1/T. The red circles correspond to data collected at a Larmor frequency ν_L= 29.49 MHz and the green triangles correspond to ν_L = 13.87 MHz. The high-temperature data from both high- and low-field experiments can be fit well to the Kubo–Tomita fit (dotted lines), both revealing an activation energy, Ea = 0.185 kcal·mol^{-1}, and a preexponential factor, $\tau_0^{-1} = 4.7 \times 10^{10}$ s^{-1}.

With respect to the conformational twisting mode of the BCO rotator, force-field calculations reported by Dunitz suggested a D_3h transition state that is only ~0.1 kcal mol^{-1} higher

than two enantiomeric D_3 structures, but likely to have a higher zero-point energy level that explains the mirror-symmetric D_3h structure inferred by single-crystal X-ray diffraction [20]. When this was explored with modern computational tools, we found that the detailed results depend on the method, but lead to the same key conclusions within the context of our study. While the MP2/6–31G* method and the B3LYP/6–31G(d) hybrid functional suggest that the D_3h conformation is a minimum with oscillations that explore the twisted D_3^+ and D_3^- structures (Fig. 4A), the more advanced M062X/6– 311+G(d) and ωB97xd/6–311+G(d) hybrid functionals find the D3h structure to be 0.2–0.35 kcal mol^{-1} higher than the twisted D_3^+ or D_3^- enantiomers. It should be noted that these small values fall within the limits of current computational methods and should be considered qualitative. Furthermore, they are comparable with the zero-point energy of the corresponding oscillation (0.3 kcal·mol^{-1}, or 113 cm^{-1}), suggesting that the enantiomerization process is more a twisting mode than an equilibrium process between two discrete energy minima. With respect to the effect of BODCA-MOF dynamics, it should be noted that the magnetic modulation caused by normal-mode frequencies of ~1012 s^{-1} is too far from the Larmor frequency of ^1H NMR in most spectrometers (107–109 s^{-1}), and is not sufficiently temperature dependent, such that this structural twisting cannot contribute to the spin-lattice relaxation. By contrast, rotational motion depends on temperature-dependent, large-amplitude angular displacements about the BCO1,4-axis expected to occur with a very short correlation time τ_c, which can be slowed down to the regime where the frequency $1/\tau_c$ approaches the Larmor frequency of the ^1H or ^2H nuclei.

The solid-state ^1H NMR T_1 measurements were carried out in the range of 2.3–80 K using a homemade cryoprobe with the sample sealed in a capillary tube under ultrahigh-purity helium. Experiments were carried out at spectrometer frequencies of 13.87 and 29.49 MHz. As indicated

in Fig. 4B with semilogarithmic plots of the relaxation rate ($1/T_1$) versus inverse temperature (1/T), there is a well-defined dynamic process with components near the Larmor frequency of 1H that occurs in the neighborhood of 20K at both spectrometer fields. There is a small and relatively constant relaxation rate at the lowest temperatures consistent with lattice modes with a very low spectral density in the megahertz regime. As the temperature increases above ~6 K (1/T= 0.16) the spin-lattice relaxation rate increases, suggesting the activation of dynamic processes that are approaching the spectrometer frequency. A maximum is observed at ~16.0K (1/T = 0.0625) when the Larmor frequency is 13.87 MHz (Fig. 4B, greentriangles), and another one at18.2K (1/T =0.0549) when the Larmor frequency increases to 29.49 MHz (Fig. 4B, red circles). Increasing the temperature to 80 K in both cases causes a linear decrease in the relaxation rate as the dynamics of the rotor exceed the Larmor frequency. Application of Eq. 2 [16] to the data acquired at 13.87 MHz for $T \geq 16$ K reveals an activation energy of 0.185 kcal·mol^{-1}. A similar analysis of the data acquired at 29.49 MHz for $T \geq 18.2$ K discloses the same activation energy. The preexponential factor required by the two fits is given by the rotational correlation time (τ_0) that would be observed in the absence of a barrier, which corresponds to a frequency factor (or attempt frequency), τ_0^{-1} = 4.7 × 1010 s^{-1}. This value is an order of magnitude smaller than the gas-phase inertial rotation estimated from the moment of inertia of the BCO group along its1,4-axis [21], τ_0^{-1} = 3.9 × 1011 s^{-1}, or it can be associated with a low-frequency torsional mode (ν_T) of 1.5 cm^{-1} that is smaller than the value ν_T =113cm^{-1} obtained by density-functional theory calculations with bicyclo[2.2.2] octane dicarboxylate.

Temperature-Dependent Rotational Dynamics and 2H LineShape Analysis

To confirm that the spin-lattice relaxation shown in Fig. 4B occurs primarily by a rotational mechanism, we carried out variable-temperature quadrupolar echo solid-state ^{2}H NMR

measurements with deuterium-labeled samples of BODCA-dMOF in the range of 292 to 6 K (Fig. 5A). The method relies on spectral changes that occur as a result of the dynamic modulation of the quadrupolar interaction between the nuclear spin and the electric field gradient at the nucleus, which depends strongly on the orientation of the C–D bond vector with respect to the direction of the external magnetic field [22–24]. Powdered samples with hydrocarbon deuterons that are either static, or undergoing siteexchangeprocessesslowerthan ∼103 s^{-1}, are characterized by a broad symmetric spectrum known as a Pake pattern that consists of two singularities and two outer shoulders with a separation that depends on the magnitude of the quadrupolar coupling constant (QCC). In the case of BODCA-d, a quadrupolar coupling constant of 168 KHz, consistent with those determined for other rigid saturated hydrocarbons, was determined from the distance between the two peaks ($\Delta\nu = 3/4$ QCC) [25] in the spectrum measured in the lowest-temperature (6 K) spectrum (see below). It is well known that the spectrum of solid samples with groups that experience axial rotation changes as a function of (i) the cone angle formed between the C–D bond vector and the rotational axis, (ii) the number of sites determined by their axial symmetry order, (iii) their corresponding angular displacements, and (iv) their rotational frequencies.In the case of the BCO group, one can observe spectra determined by C–D bonds that explore six- and threefold rotational trajectories with a cone angle of ∼70° determined by an average D_3h symmetry resulting from rapid enantiomerization. Importantly, this method has the potential to distinguish between well-defined discrete jumps among adjacent energy minima in the range of 104 to 108 Hz, and diffusional motion with less-defined angular displacements, which is expected when activation barriers are very low. [26]

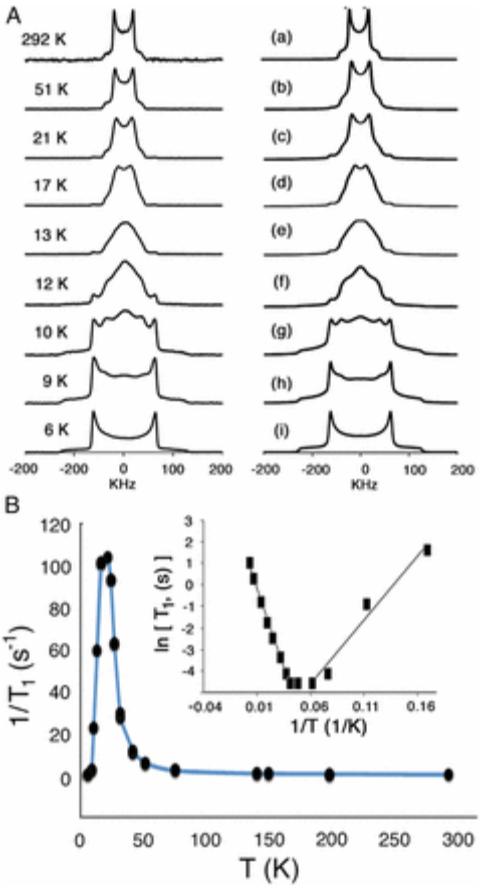


Figure 8.3. (A) ^{2}H NMR experimental (Left) and simulated spectra (Right) of crystalline BODCA-d-MOF obtained between 292 K and T ≥ 50 K. Spectra can be simulated with a model that considers either threefold or sixfold rotational motion in the fast exchange or diffusional limit. The data recorded between 50 and 21 K can be simulated (A, b and A, c, respectively) by considering motion in the fast exchange in a threefold potential with a Gaussian distribution of jump angles. The same model, but with decreasing exchange frequencies, accounts for data between 17 and 9 K (simulations A, d–A, h). The spectrum obtained at 6 K is characteristic of the slow-exchange regime with rotation lower than ∼103 s^{-1} (simulation A, i). (B) Zeeman spin-lattice relaxation for the ^{2}H nuclei in BODCA-d-MOF at 76.78 MHz illustrating a single temperature-dependent dynamic process with an optimal relaxation at 21 K. (Inset) Arrhenius plots of the data in the range of 6 to 100 K indicate activation energies of 0.12 and 0.29 kcal·mol^{-1}, respectively, for the low- and high-temperature regions.

To determine and analyze the solid-state ^{2}H NMR spectra, we were guided by the results available from the ^{1}H T_1 data. The corresponding activation parameters (Ea = 0.185 kcal·mol^{-1} and A = 4.7 × 1010 s^{-1}) indicate a modulation of magnetic interactions that will be sensitive to the ^{2}H NMR line-shape analysis method only within a relatively narrow range of temperatures from

~6 to 14 K, as the site exchange frequency within that window should vary between ~1 kHz and 10 MHz. One may predict that spectra will not change too much above this temperature range. This expectation was confirmed with measurements carried out with microcrystalline samples of BODCA-d-MOF sealed under an ultra high purity helium atmosphere in a 5-mm-diameter glass NMR tube. The spectra in Fig.5A were acquired using aquadrupolar spin-echopulse sequence at temperatures ranging from 6 to 292 K at 76.78 MHz using a home-built cryogenic 5-mm ^{2}H solid-state NMR probe at Pacific Northwest National Laboratory Environmental Molecular Science's High Field Magnetic Resonance Facility (further experimental details may be found in SI Appendix). Spectra acquired between 292 and ~50 K are very narrow and consistent with the averaging expected from unrestricted diffusional motion. As expected from the structural model, the distance between the two central peaks is simulated well by acone angle of 70° betweenthe C–D bond vector and the BODCA rotational axis. The spectra measured in the range of 50 K > T > 12 K remained very narrow, but lowering the temperatures resulted in the broadening of the two central peaks followed by their coalescence at 13 K. The low-temperature spectra were also characterized by the appearance of a small but growing contribution of a static component, suggesting that the sample is heterogeneous with sites that have slower dynamics. The line-shape changes observed down to 17 K can be simulated with a model that involves motion in the fast exchange regime (>107 s^{-1}) using mean angular displacements of 60° with a Gaussian distribution of ±15°, which are indicative of a smooth transition between diffusional motion and a site exchange process. Simulations of the spectra measured between 13 and 9 K required a reduction of rotational symmetry from six-fold to three-fold as the exchange frequency enters the intermediate exchange regime, and with an increasing contribution of the static component. Also in agreement with the

^{1}H T_1 data, it is not until the temperature reaches 6 K that the spectrum displays the characteristic pattern of the slow exchange regime.

Temperature-Dependent Solid-State ^{2}H NMR Spin-Lattice Relaxation

Finally, to demonstrate that the dynamic behavior of the ^{2}H-labeled sample is essentially identical to that of the natural abundance BODCA-MOF, we carried out a set of solid-state ^{2}H T_1 measurements. Although the rules that govern the spin lattice relaxation of quadrupolar nuclei are slightly different, an optimal relaxation at ∼21 K also indicates a component of the rotational motion that averages the Zeeman interaction at frequency of $2\pi \times 76$ MHz. A simple analysis of the high- and low-temperature regions of the Zeeman ^{2}H T_1 data shown in Fig. 5 B (Inset) corresponds to activation energies of ∼0.12 and 0.29 kcal·mol^{-1}, respectively, which is in good agreement with the value 0.181 derived from the solid-state ^{1}H T_1 data in Fig. 4B.

Molecular Dynamics Simulations

Searching for additional insights into the rotational dynamics of BODCA-MOF, we carried out molecular dynamics (MD) simulations using a model that consists of a BODCA molecule with the terminal carboxyl group fixed. As a starting point, classical MD simulations were carried out at 6, 41, and 292 K with single point charges computed with HF/6–31G(d) theory. All quantummechanical (QM) calculations were performed using G09 and the antechamber module in Amber12 to process atomic partial charges from the QM computations. Atomic partial charges were assigned with the RESP model, and the generalized Amber force field was used for further classical MD simulations. The BODCA structures were first optimized in the gas phase, heated to 6, 31, 41, or 292 K, and allowed to equilibrate for 5 ns before the corresponding MD trajectories were generated and collected for 10 ns each. Illustrated in Fig. 6 A–C are the positions of the BODCA rotor defined by the dihedral angle of O–C1–C2–C3 from 360° to −360° as a function of

time at 6, 31, and 41 K, respectively. An MD run at 298 K was virtually indistinguishable from the one at 41 K. Fig. 6D shows changes in the enantiomerization angle showing oscillations around the symmetric D_3h structure that take place at all temperatures in the picosecond time scale, although the amplitude of the displacements variesfrom ∼±5° at 6K,to ±10° at 40K,to ±20° at 298K.One can see in Fig.6A that discrete Brownian jumps of ±60° occur only at the lowest temperature while inertial or diffusional motion takes place at 31and 41K.While discrete jumps are characterized by a distribution of residence times ranging from a few picoseconds to ∼1 ns, diffusional motion in Fig.6 Band C is characterized by jumps with a wide variation in angular displacement with the rotator spending no time at any given site. It should benoted that the dynamics simulations in Fig. 6 are based on a potential that effectively underestimates the energy barrier, which results in 6-K Brownian jumps with a frequency on the order of 2 ns^{-1}, compared with∼200 μs^{-1} that would be consistent with the experimental data. It is clear, however, that while these discrepancies lead to substantial differences at cryogenic temperatures, they are much smaller at 41 K and above. Furthermore, the qualitative consistency among the experimental and computational results helps establish with great confidence that ambient temperature rotation of the BCO group in BODCA MOF occurs in the diffusional regime and constitutes an example where engineered rotational dynamics in the solid state are as fast as they would be in a high-density gas or in a low-density liquid phase.

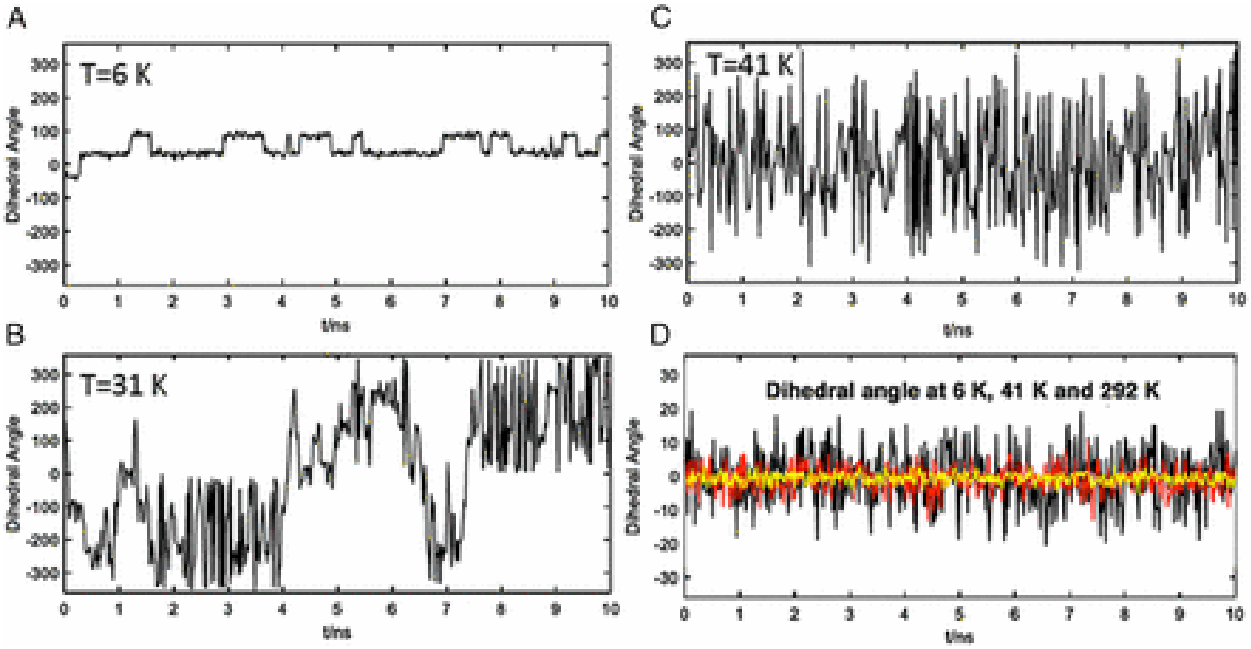


Figure 8.6. Dynamic simulations for rotation of crystalline BODCA carried out for 10-ns windows at (A) 6 K, (B) 31 K, and (C) 41 K. The amplitudes of the enantiomerization process for runs at 6 K (yellow), 41 K (red), and 292 K (black) are shown in D.

8.5 References

↵1. Kottas GS, Clarke LI, Horinek D, Michl J (2005) Artificial molecular rotors. Chem Rev 105:1281–1376.

2. Michl J, Sykes EC (2009) Molecular rotors and motors: Recent advances and future challenges. ACS Nano 3:1042–1048.

3. Vogelsberg CS, Garcia-Garibay MA (2012) Crystalline molecular machines: Function, phase order, dimensionality, and composition. Chem Soc Rev 41:1892–1910.

4. Thomas JC, et al. (2015) Defect-tolerant aligned dipoles within two-dimensional lattices. ACS Nano 9:4734–4742.

5. Horinek D, Michl J (2005) Surface-mounted altitudinal molecular rotors in alternating electric field: Single-molecule parametric oscillator molecular dynamics. Proc Natl Acad Sci USA 102:14175–14180.

6. Lemouchi C, et al. (2011) Ultra-fast rotors for molecular machines and functional materials via halogen bonding: Crystals of 1,4-bis(iodoethynyl)bicyclo[2.2.2]octane with distinct gigahertz rotation at two sites. J Am Chem Soc 133:6371–6379.

7. Rodríguez-Molina B, Pérez-Estrada S, Garcia-Garibay MA (2013) Amphidynamic crystals of a steroidal bicyclo[2.2.2]octane rotor: A high symmetry group that rotates faster than smaller methyl and methoxy groups. J Am Chem Soc 135:10388–10395.

8. Catalano L, et al. (2015) Dynamic characterization of crystalline supramolecular rotors assembled through halogen bonding. J Am Chem Soc 137:15386–15389.

9. Gould SL, Tranchemontagne D, Yaghi OM, Garcia-Garibay MA (2008) Amphidynamic character of crystalline MOF-5: Rotational dynamics of terephthalate phenylenes in a free-volume, sterically unhindered environment. J Am Chem Soc 130:3246–3247.

10. Horike S, et al. (2006) Dynamic motion of building blocks in porous coordination polymers. Angew Chem Int Ed Engl 45:7226–7230.

11. Eddaoudi M, et al. (2002) Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. Science 295:469–472.

12. Eddaoudi M, et al. (2001) Modular chemistry: Secondary building units as a basis for the design of highly porous and robust metal-organic carboxylate frameworks. Acc Chem Res 34:319–330.

13. Huang-Minlon NA (1946) A simple modification of the Wolff-Kishner reaction. J Am Chem Soc 68:2487–2488.

14. Walton KS, Snurr RQ (2007) Applicability of the BET method for determining surface areas of microporous metal-organic frameworks. J Am Chem Soc 129:8552–8556.

15. Li H, Eddaoudi M, O'Keeffe M, Yaghi OM (1999) Design and synthesis of an exceptionally stable and highly porous metal-organic framework. Nature 402:276–279.

16. Larson AC, VonDreele RB (2004) General Structure Analysis System (GSAS) (Los Alamos National Laboratory, Los Alamos, NM), Technical Report LAUR 86-748.

17. Fyfe CA (1983) Solid-State NMR for Chemists (CFC Press, Guelph, Ontario).

18. Kubo R, Tomita KA (1954) General theory of magnetic resonance absorption. Phys Soc Jpn 9:888–919.

19. Bloembergen N, Purcell EM, Pound RV (1948) Relaxation effects in nuclear magnetic resonance absorption. Phys Rev 73:679–712.

20. Ermer O, Dunitz JD (1969) Zur konformation des bicyclo[2.2.2]octan-systems. Helv Chim Acta 51:1861–1885.

21. Jiang X, Rodríguez-Molina B, Nazarian N, Garcia-Garibay MA (2014) Rotation of a bulky triptycene in the solid state: Toward engineered nanoscale artificial molecular machines. J Am Chem Soc 136:8871–8874.

22. Spiess HW (1983) Molecular dynamics of solid polymers as revealed by deuteron NMR. Coll Polym Sci 261:193–209.

23. Hoatson GL, Vold RL (1994) 2H NMR spectroscopy of solids and liquid crystals. NMR Basic Princ Prog 32:1–67.

24. Luz Z, Hewitt RC, Meiboom S (1974) Deuterium magnetic resonance study of a smectic liquid crystal. J Chem Phys 61:1758–1765.

25. Rice DM, et al. (1981) Rotational jumps of the tyrosine side chain in crystalline enkephalin. Hydrogen-2 NMR line shapes for aromatic ring motions in solids. J Am Chem Soc 103:7707–7710.

26. Vugmeyster L, et al. (2009) Probing the dynamics of a protein hydrophobic core by deuteron solid-state nuclear magnetic resonance spectroscopy. J Am Chem Soc 131:13651–13658.

Chapter 9. Molecular Dynamics Study of Gearing and Slippage Motions in Molecular Spur Gears

9.1 Abstract

Inspired by biological molecular machines, such as ATP synthase and motor proteins, a variety of artificial molecular machines have been developed since 1980s. Among the functional molecular machines, molecular gears involve two or more molecular rotors in contact so that they are able to transfer motions through geared rotation. Two modes of rotation in molecular gears are possible: gearing (dis-rotation or rotation in opponent directions) and slippage (con-rotation or rotation in the same direction). High gearing efficiency (gearing-to-slippage ratio) is desirable for both macroscopic gears and molecular gears as it can transmit motions efficiently. Molecular bevel gears have been developed by Mislow and Iwamura and achieved high gearing efficiency. However, the most common gears in the macroscopic world, spur gears, are less well established at the molecular level. The designed and synthesized molecular spur gears have not yet achieved efficient gearing. In this work, we studied the rotation of molecular spur gears using molecular dynamics (MD) simulations in the gas phase and in organic solvents. The potential of mean force (PMF) of both slippage and gearing processes were computed using 2D metadynamics. We discovered several parameters which have great influence on the gearing efficiency. Most importantly, we found that organic solvent molecules promote slippages of molecular spur gears by stabilizing the slippage transition state. Furthermore, we proposed a promising molecular spur gear which shows high gearing efficiency in both the gas phase and organic solvents as tested in MD simulations.

9.1 Introduction

In the macroscopic world, machines often involve motion transmission by gears composed of two or more rotors that move together, the rotation of one drives the rotation of the others. The prototype biological machine, F1-ATP synthase, consists of two rotary motors Fo and F1. Under physiological conditions, the two rotors gear and lead to ATP synthesis or generation of a proton motive force (pmf).[1,2] Well-functioning molecular gears are also desirable for the development of artificial molecular machines. The initial study of molecular gears can be dated back to the 1980s when Mislow and Iwamura independently studied dynamic gearing in molecular bevel gears.[3–6] They synthesized the molecular bevel gears Tp2X consisting of two 9-triptycyl (Tp) groups joined by the bridgehead carbons to a central atom X (C or O). The bevel gear systems achieved high efficiency towards gearing, with the energy barrier of correlated dis-rotation (geared rotation) to be only 1-2 kcal/mol. In comparison, the energy barriers for correlated con-rotation (slippage) are in the range of 32-45 kcal/mol. More recently, Shionoya et al. developed molecular bevel gears with multiple rotors and molecular bevel gears with photo- and thermal- driven mechanical switching.[7–9] However, the development of molecular spur gears with high gearing efficiency, which is the most common gear in macroscopic machinery, is more difficult than molecular bevel gears. The success of molecular bevel gears originates in the fact that the rotors are tightly intermeshed. In the case of Tp2X, the spatial arrangement of the rotors causes a high penalty to slippage, since the clash of two phenyl rings and the distortion of the Tp-X-Tp bond angles require extremely high energy. However, in molecular spur gears, it is difficult to achieve high gearing efficiency as rationalized by Siegel et al.[10] Geared rotation requires more separation of the core axes of the Tp groups than gear slippage.

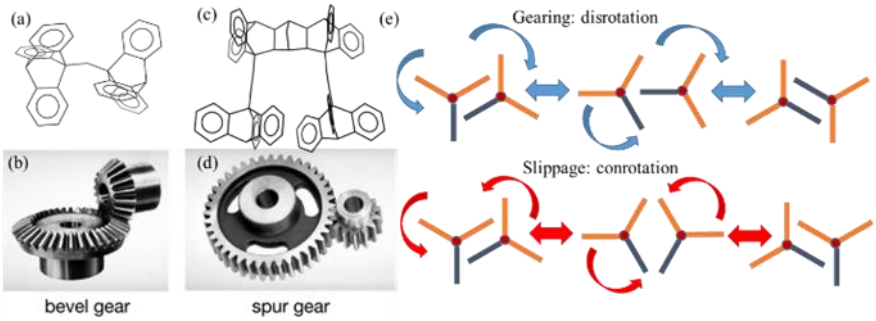


Figure 9.1. Illustration of the molecular gears and their corresponding macroscopic systems: (a) molecular and (b) macroscopic bevel gears; (c) molecular and (d) macroscopic spur gears; (e) rotations of two rotors in a gear system, gearing is dis-rotation of the two rotors while slippage happens as the con-rotation of the two rotors.

Existing molecular spur gears are composed of three parts, stators, rotors and axles to connect them (Figure 2). The Wakamatsu group synthesized a molecular spur gear with triptycene rotors attached to an acene base.[11,12] They predicted that the geared rotation is preferred by 7 kJ/mol compared to gearing slippage using hybrid DFT M05/6-31G. However, signal broadening in the 1HNMR spectra was observed and is attributable to the slippage motions. Siegel et al. proposed a hard-sphere model to guide the design of highly efficient spur gears and concluded the inter-axle distance is the key to design molecular spur gears. An inter-axle distance of roughly 8 Å is optimal for high gearing efficiency over slippage.[10] Later they developed and synthesized a spur gear system based on their model, which was predicted to prefer gearing over slippage by 4.2 kcal/mol using DFT calculations.[13] However, no decoalescence in VT-NMR experiment was observed. The absence of decoalescene indicated that the gearing in their system was not dominant enough to be observed in VT-NMR experiment. It suggests that the inter-axle distance is not the only parameter that affects the behavior of molecular spur gears.

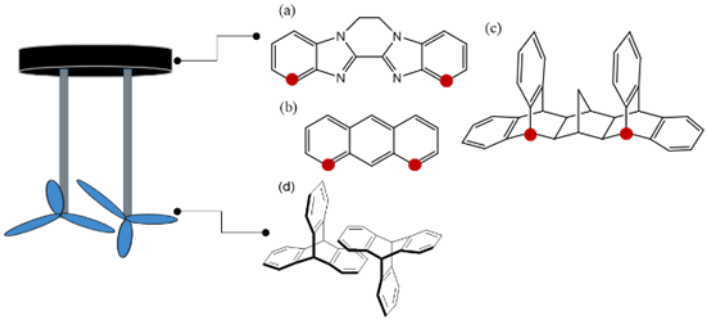


Figure 9.2. Molecular spur gear systems are composed of three parts: stator (black), axle (grey) and the rotor (blue); (a) Garcia-Garibay stator, red points are where axles attach; (b) Wakamatsu stator; (c) Siegel stator; (d) two triptycenes act as the rotors in all molecular spur gears to date.

Therefore, we systematically investigated the molecular spur gears developed by different research groups, using methods from molecular dynamics (MD) and steered MD simulations (metadynamics) to DFT calculations. The influence of the stator structure and the inter-axle distance are investigated in this study. Moreover, as the VT-NMR experiments were often carried out in organic solvents, we also looked into the influence from the surrounding organic solvent molecules on gearing. The inter-axle distances are found to have a great impact on the rotational motions, as illustrated in Siegel's model. Inter-axle distances than 8 Å result in spur gears that mostly undergo slippage. The interaction between stator and rotors is able to shape the potential energy surface (PES) of rotational motions, biasing the system towards gearing or slippage. Most importantly, the solvent molecules, which were previously neglected, can interact with the triptycenes and promote slippage by stabilizing the slippage TS. Based on the insights from theoretical study, we designed a promising molecular spur gear that shows dominant geared rotations in the gas phase and organic solvents as tested in MD simulations.

9.2 Computational Methods

Molecular Dynamics (MD).

Molecular dynamics simulations of the molecular spur gears were performed using the GPU code (pmemd) of the Amber 14 package.[14–16] The parameters were generated with the

antechamber module using the general Amber force field (gaff)[17,18], and the atomic partial charges were assigned using the RESP method.[19] Partial charges were computed with HF/6-31G* theory. We used Gaussian09[20] for the quantum mechanical (QM) computations and the antechamber module in Amber14 to process atomic partial charges from the QM computations. In the MD simulations in the gas phase, the structures of each gear system were first minimized in vacuum with no solvent model and then heated from 0 to 300 K gently under a constant pressure of 1 atm. The molecular gears were then equilibrated for 2 ns with 1 fs time steps at a constant volume. Production trajectories were then run for additional 1 ns under the same simulation conditions. The MD simulations in organic solvents were carried out with the same protocol except that the gear molecules were immersed in an organic solvent box. The dihedral angles of two triptycene groups in each gear system were tracked for further data analysis.

Metadynamics.

Metadynamics MD simulations were performed using the NAMD v2.11 program.[21,22] The input files were generated using the same method as in MD simulations. Well-tempered metadynamics in conjunction with a multiple-walker algorithm were used to accelerate the convergence of the free energy landscape computed from the metadynamics simulations. The Langevin thermostat controlled the temperature at 300 K and Langevin barostat kept the pressure at 1 atm. The collective variables were chosen as the dihedral angle of the two triptycenes of each molecular gear, with the range constrained to be -180° to 180º. The gear systems were first minimized and equilibrated for 20 ns. We simulated the rotations of each spur gear for 180 ns, and 2D potentials of mean force were generated.

DFT Calculations.

All density functional theory calculations were performed using Gaussian09.[20] Geometry optimizations and subsequent frequency calculations were performed at the B3LYP/6-31G(d) level of theory. Normal vibrational mode analysis confirmed that optimized structures were ground states (no imaginary frequency) or transition states (one imaginary frequency). Single-point energies were computed at the B3LYP/6-311+G(d,p) level of theory. Reported energies are Gibbs free energies determined by summing the single-point electronic energies and thermal corrections determined using B3LYP/6-31G(d).

9.3 Results and Discussions

9.3.1 Molecular Dynamics and Metadynamics.

We studied the gearing behavior of different molecular gears (Figure 3) in the gas phase with classical MD simulations. 1 is the molecular bevel gear developed by Mislow and Iwamura, while all the rest are molecular spur gears. 2 and 4 are molecular spur gears proposed and synthesized by Wakamatsu and Siegel, respectively. 3 was developed in the Garcia-Garibay group at UCLA. All molecular gears have two triptycenes as the rotors. The molecular spur gears have different stators, which result in different inter-axle distances. In 2 and 3, the inter-axle distances are relatively smaller, which leads to a more compacted arrangement of the two rotors than structure 4.

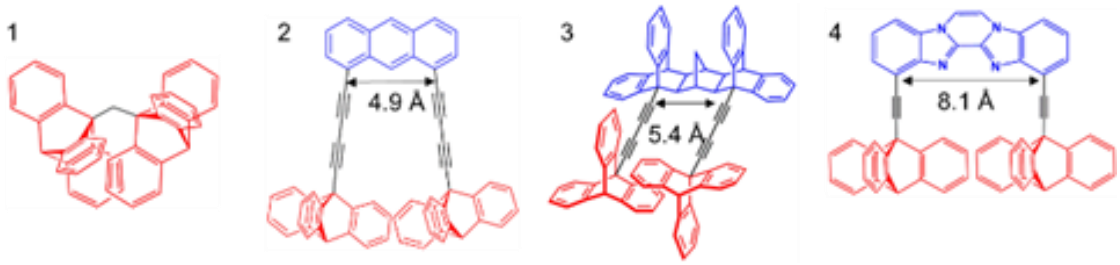


Figure 9.3. Molecular gear systems studied with classical MD and metadynamics. System 1 is a bevel gear developed by Mislow and Iwamura; systems 2-4 are spur gears with different stator.

MD simulations are able to reveal the dynamic information of the rotational processes for all molecular gears. Comparing the rotational directions of the two rotors in molecular gears, we can easily determine whether the molecular gears go through gearing or slippage at a certain time point. To track the rotation of the two rotors, the dihedral angles between the rotor and the stator are extracted from the MD trajectories. Figure 4 illustrates how the dihedral angles change in the gearing and slippage processes of the molecular gears. In the gearing process, one dihedral angle decreases and the other dihedral angle increases as the two rotors dis-rotate. In the slippage process, the dihedral angles changes are the same, because the two rotors con-rotate.

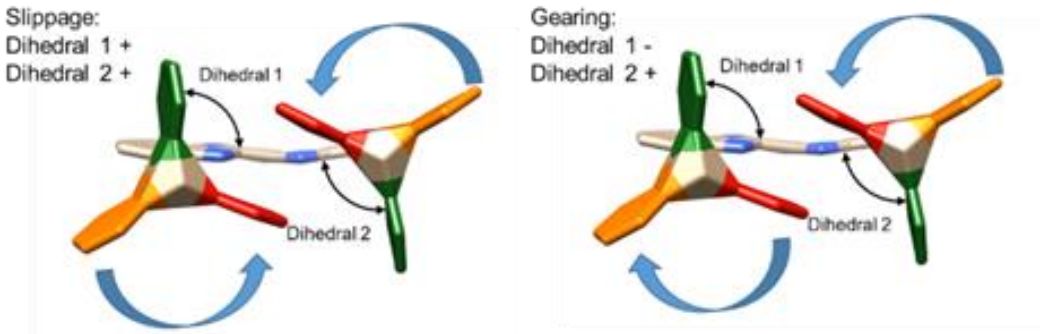


Figure 9.4. Dihedral angle changes in the slippage (left) and gearing (right) processes.

Figure 5 shows the dihedral plots for all four molecular gears. In molecular spur gear 1, the dihedral plot is highly symmetric and the two dihedral angle changes are opposite. This indicates that the rotations in the molecular bevel gear are correlated and the two rotors dis-rotate. The molecular spur gear 1 goes through only gearing during the 1 ns MD simulation. In molecular spur gears 2-3, we observed both same dihedral changes and opposite dihedral changes in the dihedral angle plot. The rotations of the two rotors in 2-3 are also correlated. However, both gearing and slippage are observed for 2-3 during the 10 ns MD simulation. In 4, the dihedral angle changes are mostly opposite, which indicates that 4 is a good molecular gear with a high gearing efficiency in the gas phase.

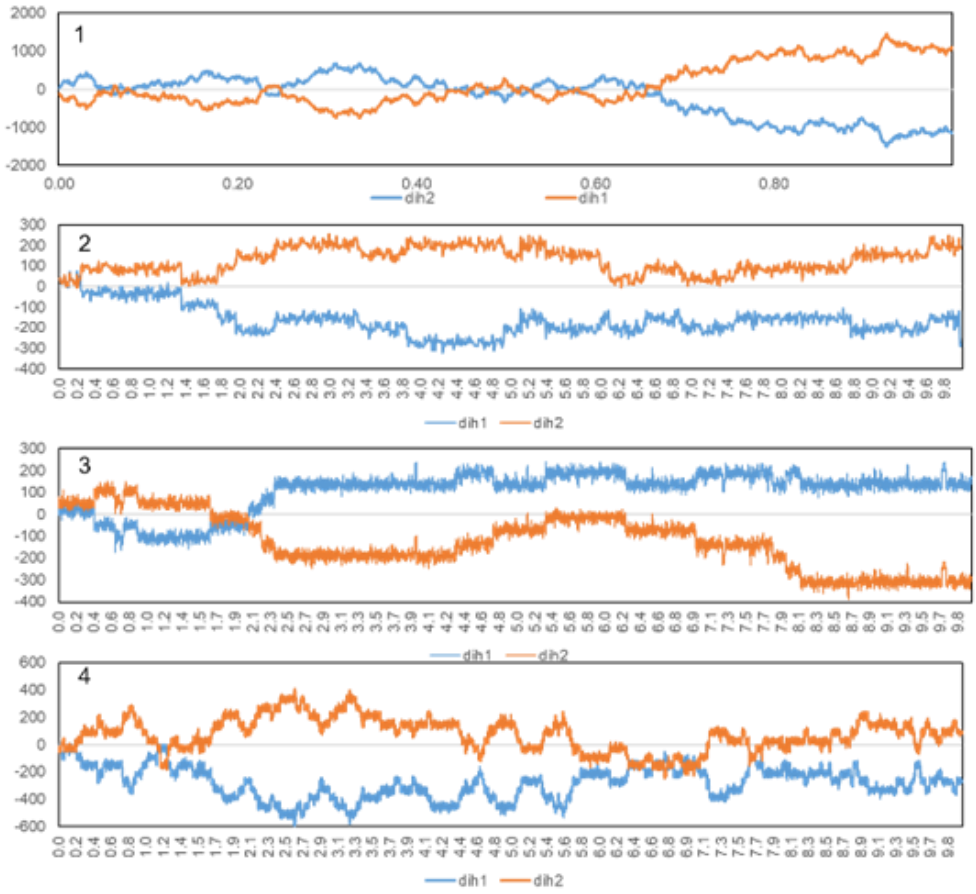


Figure 9.5. Dihedral angle changes for two rotors in systems 1-4.

The classical MD simulation results are consistent with the hard-sphere model proposed by Siegel et al. The average inter-axle distance (from the center of a rotor to the center of the other) is 6.67 Å for 2 and 7.0 Å for 3. Both of the distances are smaller than 8 Å, the minimum inter-axle distance for the gearing transition state (TS). In comparison, 4 has an average inter-axle distance of 8.1 Å, which is slightly larger than the required distance. However, at such a distance, we suspect the slippage is also possible and only slightly unfavorable compared to gearing.

2D Metadynamics was performed to compute the free energy surface (in terms of potential mean force PMF) of the rotations of the two rotors. We chose the dihedral angles (defined in Figure 4) corresponding to the rotation of the triptycenes as the collective variables and scanned the range from -180º to 180º for each collective variable. In the 2D PMF plot of rotations (Figure 6), the

diagonals represent gearing (red line) and slippage (black line). The bevel gear 1, is the most efficient molecular gear with a gearing energy barrier 1-2 kcal/mol and the slippage energy barrier to be over 27 kcal/mol. For gears 2-3, the energy barriers of slippage are about 1-2 kcal/mol lower than gearing, which is consistent with classical MD simulations. In 4, the energy barrier of gearing is about 1 kcal/mol lower than slippage. Though the energy difference between gearing and slippage for 4 is much smaller than what Siegel reported (3.4 kcal/mol) using the DFT method, we also concluded that gearing is more favorable than slippage for 4.

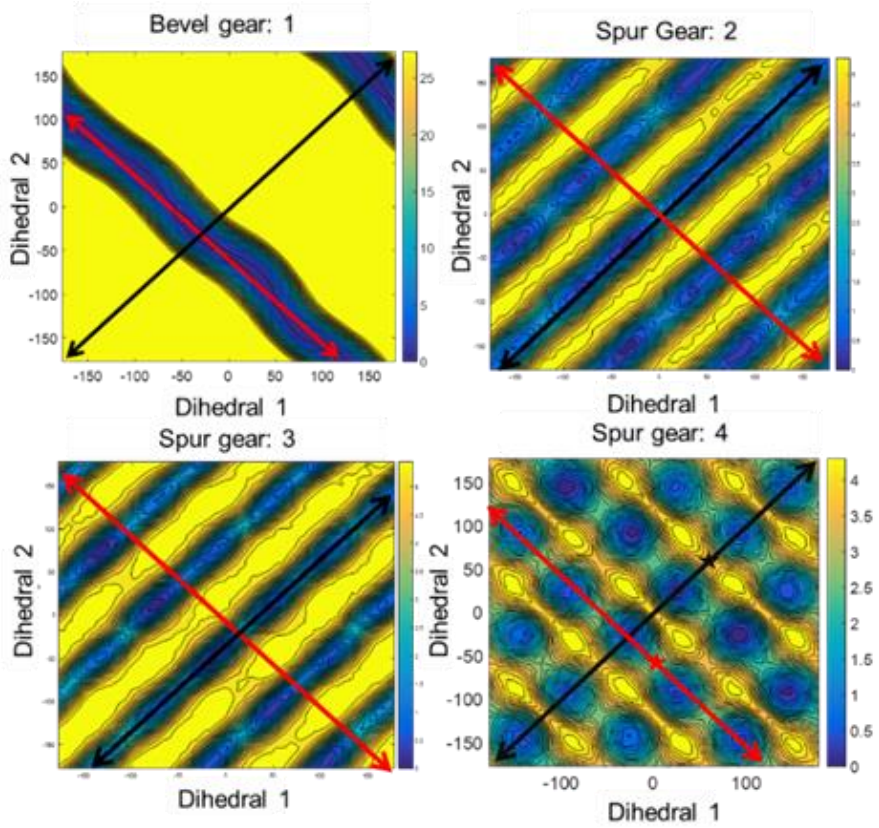


Figure 9.6. 2D PMF from metadynamics. Collective variables are the dihedral angles of C (stator) –C (axle) –C (axle) –C (triptycene) as labeled as red dots. Metadynamics simulations scanned the collective variables (dihedral angle 1 and 2) from -180º to 180º.

9.3.2 Influence of Organic Solvents.

To better understand the dynamics of molecular spur gears in the VT-NMR experiments, we carried out classical MD simulations of 4 in organic solvents (toluene and $CHCl_3$). Figure 7 shows that dihedral plot of system 4 in the gas phase, toluene and $CHCl_3$. In the gas phase, the dihedral plot is highly symmetric which indicate that the two rotors go through mostly geared rotation, as the dihedral angle changes are mostly opposite. However, in toluene and $CHCl_3$, the dihedral-angle plots become more asymmetric which indicates that the two rotors go through mainly slippage or uncorrelated rotation. We looked into the structure of 4 in MD simulations and found that solvent molecules can go into the space between two rotors. The solvent molecule in between the two rotors stabilizes the slippage TS of 4. With the surrounding solvent molecules included in MD simulations, the molecular spur gear 4 behaved largely different than in the gas phase.

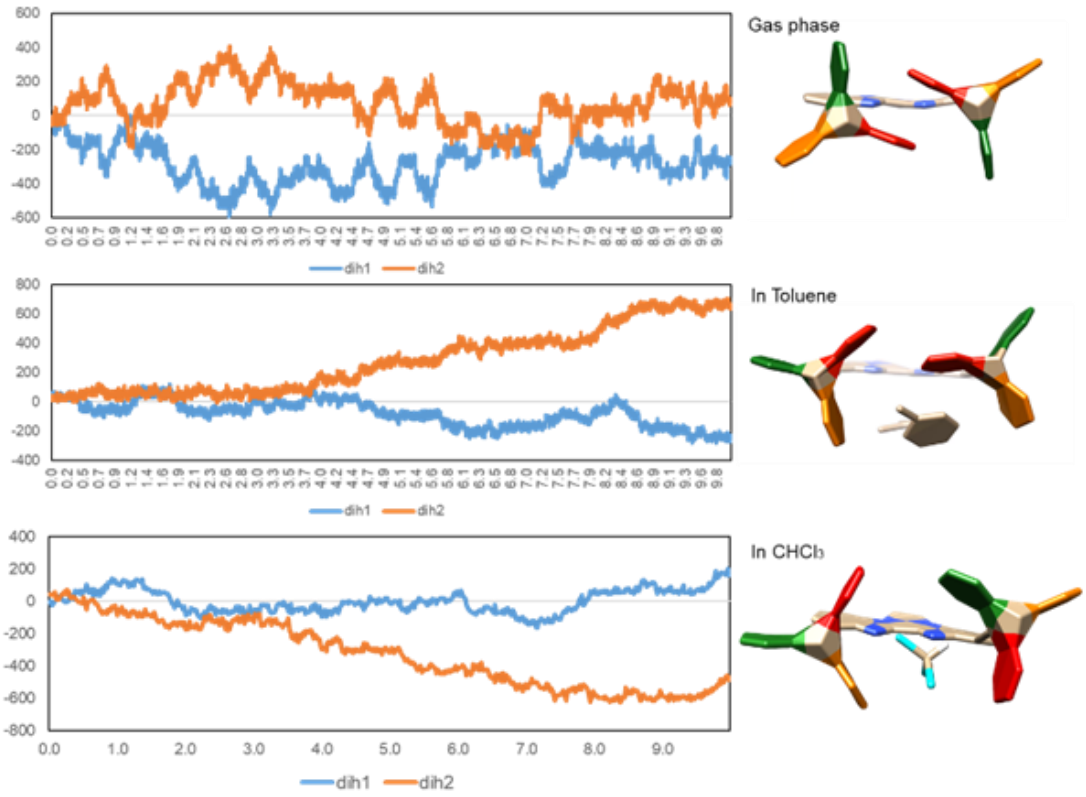


Figure 9.4. Dihedral plot of system 4 in different media.

To reduce the solvent effect on the gearing of 4, we proposed a macrocyclic molecular spur gear 5 based on 4. The macrocyclic structure of 5 is made by attaching another stator from the other end of the rotors. In the design of 5, the freedom of two rotors is reduced as the distance between the two rotors is now fixed by two stators. The limited freedom of the two rotors further prevents solvent molecules from getting into the space between the two rotors. We tested the gearing behavior of 5 in toluene and $CHCl_3$ solvents. As shown in Figure 8, the dihedral plots are symmetrical for both solvents. The dihedral changes are in opposite directions for each rotation in the MD simulations.

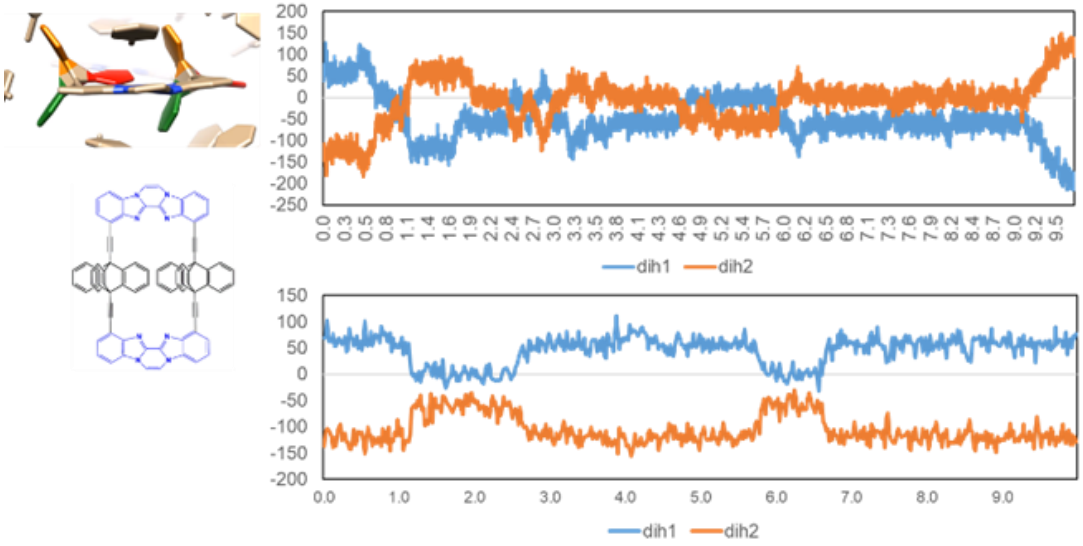


Figure 9.8. Proposed locked spur gear and the dihedral plot in toluene and $CHCl_3$

9.3.3 DFT calculations.

We further optimized the ground-state and transition-state geometries for 4 and 5 using density functional theory (DFT) calculations. The computed free energy barriers are shown in Figure 9. For comparison, the corresponding barriers for 3 are also included in Figure 6.

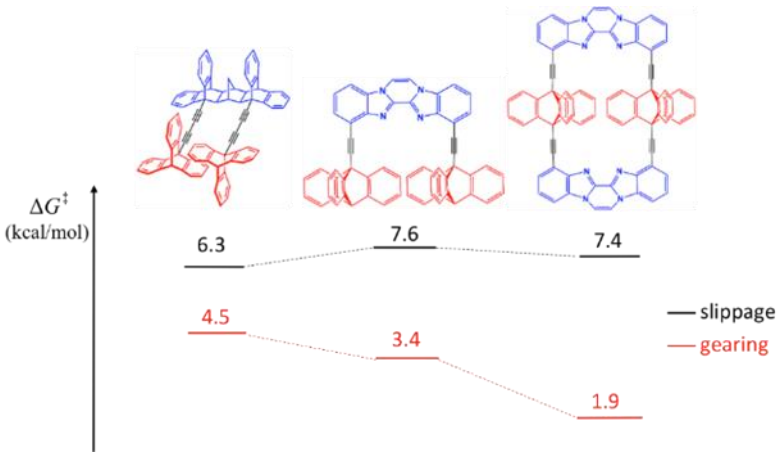


Figure 9.9. Computed free energy barriers for 3 (stator 1, half open) and 5 (stator 3, locked). Free energies were computed at the B3LYP/6-311+G(d,p)//B3LYP/6-31G(d) level. For comparison, the corresponding barriers for system 4 (stator 3, half open) are also included.

For each system, gearing has a lower barrier than slippage. The gearing of 3 has a barrier of 4.5 kcal/mol, while its slippage has a higher barrier of 6.3 kcal/mol. Thus, gearing is favored over slippage by only 1.8 kcal/mol. For 4, the barrier of gearing decreases (3.4 kcal/mol) while the barrier of slippage increases (7.6 kcal/mol). This results in a larger gap between the two transition states for 4, and gearing is now much more favored than slippage, by 4.2 kcal/mol. From the half-opened system 4 to the locked system 5, this gap becomes even larger (5.5 kcal/mol). This is mainly due to the decrease in the energy of the gearing TS (from 3.4 to 1.9 kcal/mol). The results indicate that 4 and 5 have better gearing efficiency compared to 3. To explain this observation, we examined the ground-state geometry and the transition structures for each system. C-H···N interactions were identified in the gearing TSs of 4 and 5. Figure 10 shows the gearing TSs and slippage TSs for 4 and 5, respectively.

For the gearing TS of 4, C-H···N interaction exists between one aromatic C-H and one of the N atoms on the stator (2.6 Å). This distance is longer (3.8 Å) and presumably weaker in the corresponding slippage TS. Therefore, the C-H···N interaction stabilizes the gearing TS relative to the slippage TS. In system 5, one additional C-H···N interaction is introduced as another stator

is added from the other end. This further reduces the energy barrier of gearing. From the metadynamics results, we got a much smaller difference between the energy barriers of gearing and slippage for systems 4. This deviation is likely originated from the poorer treatment of C-H···N by the force field as compared to the more accurate density functional theory calculations.

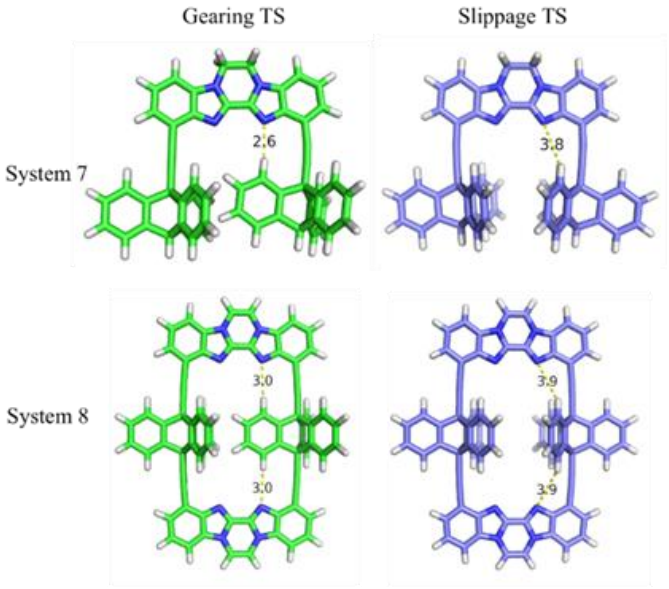


Figure 9.10. Transition structures for systems 7 and 8. C-H···N interactions are indicated with dashed lines. Distances shown are in Å.

9.4 Conclusions

Combining classical MD simulations, metadynamics and DFT calculations, we performed a systematic investigation into molecular spur gears. The inter-axle distances between the two rotors determine the arrangement of the rotors, thus are critical to the gearing behavior of the molecular spur gears. Our study compared three different molecular spur gears with different inter-axle distances and found that the inter-axle distance of about 8 Å lead to high gearing efficiency in the gas phase. However, in solvent, the organic solvent molecules are able to stabilize the slippage TS and promote slippage over gearing. We proposed a molecular spur gear with a locked macrocyclic structure which is promising to eliminate the solvent effect on gearing thus prevent

slippage even in the solvent phase. DFT calculation of molecular spur gears further reveals that the attractive interaction between rotors and stators in the gearing TS also contribute to the high gearing efficiency in our design of a molecular spur gear.

9.5 References

(1) Boyer, P. D. The ATP Synthase—A Splendid Molecular Machine. Annual Review of Biochemistry 1997, 66 (1), 717–749.

(2) Itoh, H.; Takahashi, A.; Adachi, K.; Noji, H.; Yasuda, R.; Yoshida, M.; Kinosita, K. Mechanically Driven ATP Synthesis by F1-ATPase. Nature 2004, 427 (6973), 465–468.

(3) Iwamura, H. Molecular Design of Correlated Internal Rotation. Journal of Molecular Structure 1985, 126, 401–412.

(4) Iwamura, H.; Mislow, K. Stereochemical Consequences of Dynamic Gearing. Acc. Chem. Res. 1988, 21 (4), 175–182.

(5) Kawada, Y.; Iwamura, H. Unconventional Synthesis and Conformational Flexibility of Bis(1-Triptycyl) Ether. J. Org. Chem. 1980, 45 (12), 2547–2548.

(6) Wd, H.; Ca, J.; A, G.; F, C.; K, M. Stereochemical Consequences of Dynamic Gearing in Substituted Bis(9-Triptycyl) Methanes and Related Molecules., Stereochemical Consequences of Dynamic Gearing in Substituted Bis(9-Triptycyl) Methanes and Related Molecules. Proc Natl Acad Sci U S A 1980, 77, 77 (12, 12), 6961, 6961–6964.

(7) Sanada, K.; Ube, H.; Shionoya, M. Rotational Control of a Dirhodium-Centered Supramolecular Four-Gear System by Ligand Exchange. J. Am. Chem. Soc. 2016, 138 (9), 2945–2948.

(8) Ube, H.; Yasuda, Y.; Sato, H.; Shionoya, M. Metal-Centred Azaphosphatriptycene Gear with a Photo- and Thermally Driven Mechanical Switching Function Based on Coordination Isomerism. Nature Communications 2017, 8, ncomms14296.

(9) Ube, H.; Yamada, R.; Ishida, J.; Sato, H.; Shiro, M.; Shionoya, M. A Circularly Arranged Sextuple Triptycene Gear Molecule. J. Am. Chem. Soc. 2017, 139 (46), 16470–16473.

(10) Frantz, D. K.; Baldridge, K. K.; Siegel, J. S. Application of Structural Principles to the Design of Triptycene-Based Molecular Gears with Parallel Axes. CHIMIA International Journal for Chemistry 2009, 63 (4), 201–204.

(11) Ogi, S.; Ikeda, T.; Wakabayashi, R.; Shinkai, S.; Takeuchi, M. A Bevel-Gear-Shaped Rotor Bearing a Double-Decker Porphyrin Complex. Chem. Eur. J. 2010, 16 (28), 8285–8290.

(12) Toyota, S.; Shimizu, T.; Iwanaga, T.; Wakamatsu, K. Structures and Conformational Analysis of 1,8-Bis(9-Triptycylethynyl)Anthracene and Its Derivatives as Prototypes of Molecular Spur Gears. Chem. Lett. 2011, 40 (3), 312–314.

(13) Frantz, D. K.; Linden, A.; Baldridge, K. K.; Siegel, J. S. Molecular Spur Gears Comprising Triptycene Rotators and Bibenzimidazole-Based Stators. J. Am. Chem. Soc. 2012, 134 (3), 1528–1535.

(14) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; et al. Amber 14; University of California, San Francisco, 2014.

(15) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J. Chem. Theory Comput. 2013, 9 (9), 3878–3888.

(16) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. Computer Physics Communications 2013, 184 (2), 374–380.

(17) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. Journal of Molecular Graphics and Modelling 2006, 25 (2), 247–260.

(18) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. J. Comput. Chem. 2004, 25 (9), 1157–1174.

(19) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. J. Phys. Chem. 1993, 97 (40), 10269–10280.

(20) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. Gaussian 09, Revision B.01; Gaussian, Inc.: Wallingford CT, 2009.

(21) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. J. Comput. Chem. 2005, 26 (16), 1781–1802.

(22) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. Proc Natl Acad Sci U S A 2002, 99 (20), 12562–12566.