

## Large-Scale Computations of Chemical and Biological Reaction Rates and Mechanisms

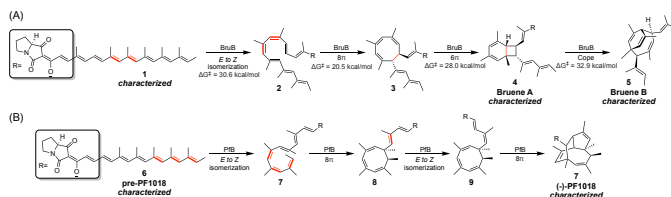
**Abstract:** This proposal seeks a renewal of computational resources for the research currently supported by ACCESS (project number [REDACTED]). We propose seven research projects involving state-of-the-art density functional theory (DFT) calculations applied to far-reaching areas of chemistry including organometallics, organocatalysis, enzyme catalysis, and mechanisms of photochemical and other organic reactions. A progress report of new scientific discoveries in 2024, with a list of publications, is also submitted.

### Science Objectives

#### *Project 1: Understanding the Mechanism of BruB and PfB Enzyme for the Biosynthesis of Polycyclic Natural Products from Conjugated Polyenes*

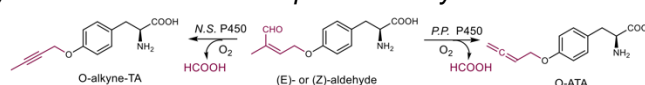
In collaboration with [REDACTED] we will conduct DFT calculations and hybrid QM/MM MD simulations to study the reaction mechanism and origin of enzyme catalysis in two novel enzymatic systems BruB and PfB.<sup>1</sup> Our initial DFT calculation and experimental results proposed a sequence of pericyclic

reactions are involved in mechanism, but the role of enzyme environment is still under investigation. We will conduct DFT and MD calculations to study origin of catalysis and stereoselectivity controlling. Considering the conformational flexibility of these enzymes, we request 500,000 core-hours to perform about 2000 geometry optimizations, frequency calculations, and single point energy computations.



#### *Project 2: Enzymatic demethylative allene and alkyne formations in natural product biosynthesis*

We are collaborating with [REDACTED] and planning to use state of the art DFT method to determine the mechanism of two interesting P450 systems. Both of them catalysis the aldehyde deformylation with the presence of O<sub>2</sub>, while different enzymes leads to different products, i.e. alkyne or allenes. Previous mechanistic studies proposed several potential pathways for similar transformation. However, the formation of allene is still novel in this area and the detailed mechanism is still unknown. For further understanding these reactions, we plan to calculate several possible pathway to predict the most feasible mechanism. Considering the complexity of iron electronic structures and flexibility of substrate conformation, we request 465,000 core-hours to perform about 800 geometry optimizations, frequency calculations, and single point energy computations.



#### *Project 3: Data-Driven Design of Electron-Deficient Olefin Ligands Enables Access to Enantioenriched Quaternary Stereocenters*

In collaboration with the [REDACTED] are employing a tandem in-silico and experimental strategy to develop a novel class of chiral electron-deficient olefin (EDO) ligands to enable the Ni catalyzed C(sp<sup>3</sup>)-C(sp<sup>3</sup>) Negishi cross-coupling of racemic 1,1-disubstituted aziridines to form enantioenriched quaternary-substituted amines.<sup>2</sup> Quantum mechanics calculations, data science techniques and experimental mechanistic studies are employed to rationalize and optimize chemo-, regio-, and enantioselectivity. Preliminary density functional theory calculations support a Ni<sup>0-II</sup> catalytic cycle, where reductive elimination is both rate-limiting and stereo-determining. We have developed a stereochemical model of this reaction for the development of new chiral ligands through a virtual library approach, and work is ongoing to experimentally evaluate the novel chiral EDO ligands developed from our computations. An estimated 800,000 CPU hours will facilitate the completion of this project.

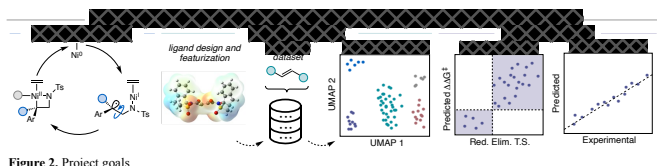
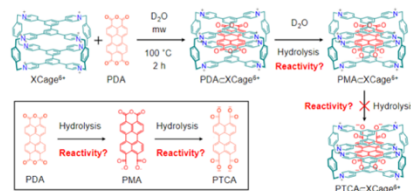


Figure 2. Project goals

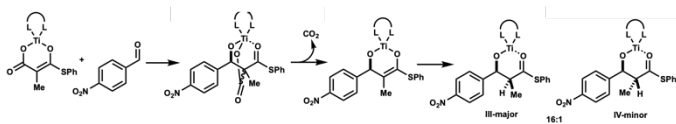
#### *Project 4: Computational Studies of Anthracene-based Molecules for Solar Energy Storage and Release*

The late Prof. J. Fraser Stoddart and Dr. Huang Wu at Hong Kong University, and Dr. Wu now at Nankai University, created an organic molecular container called XCage.<sup>3</sup> They discovered that perylene dianhydride (PDA) would hydrolysis only one of the anhydrides inside the XCage. To understand the origin of this unusual reactivity, we are using DFT method to investigate this reaction. Currently, we believe the reactivity is likely controlled by the kinetic factor, and we are working to gain deeper insights into it. Considering the size of host-guest system, to accomplish this project, we request 638,000 core hours for optimization, frequency, and high-level single-point calculations.



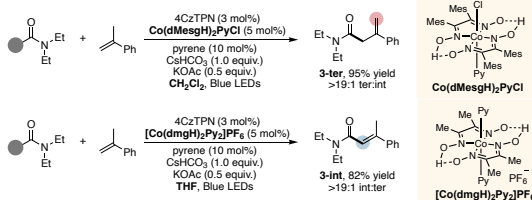
#### Project 5: Ti-Salen/Salalen Catalysis of Aldol Additions with Malonic-Acid Half Thioesters

The aldol addition is one of the most powerful carbon-carbon bond-forming reactions, but often requires the preformation of the enolate to control the stereochemistry of the resulting hydroxyketone. In our collaboration with [REDACTED], we have investigated titanium(IV) with Salen and Salalen ligands as a catalyst for the aldol addition using malonic-acid half thioesters (MAHT), which affords stereocontrol of the resulting diastereomers. We are using Density Functional Theory (DFT) in Gaussian and conformational sampling with xTB-CREST to develop an understanding of the experimentally observed selectivity. Our objective is to understand why some aldehydes lead to greater selectivity, and how changing the MAHT-methyl substituent to a fluorine or amine will affect the overall reaction. For this work, we will require 46,000 CPU hours.



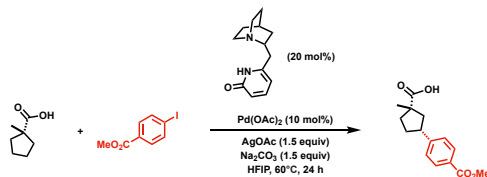
#### Project 6: Regiodivergent Olefinic C-H Carbamoylation of Branched Alkenes

In collaboration with [REDACTED], we developed a cobalt-catalyzed functionalization reaction of alkenes. The oxime-type ligand is found to be very critical in determining the regioselectivity in this transformation, namely, the thermodynamically favored products (internal alkene) were obtained when using the oxime ligand with smaller steric hindrance while the kinetically favored products (terminal alkene) were isolated when using the oxime ligand with greater steric hindrance. We apply DFT calculations to investigate this unusual regioselectivity controlled by the steric effect of ligands. Calculations will be performed at the  $\omega$ B97X-D/6-311G\*\* (SDD for Co)/SMD(DCM)// $\omega$ B97X-D/6-31G\* (SDD for Co) level of theory. For this work, we will require 200,000 CPU hours.



#### Project 7: Substrate-Controlled Regioselectivity in Transannular C-H functionalization of Cyclopentane Carboxylic Acids

In collaboration with [REDACTED], we use theoretical calculations to study a class of palladium-catalyzed transannular C-H functionalization reactions of cyclic alkane carboxylic acids. These reactions exhibit excellent regioselectivity. Our computational studies revealed the reaction mechanism and showed that the regioselectivity is governed by the intrinsic structure of the cyclopentane carboxylic acid substrates, rather than by ligand control as in traditional systems. Calculations will be performed at the PBE0-D3(BJ)/def2-TZVPP/SMD(HFIP)//PBE0-D3(BJ)/def2-SVP level of theory. For this work, we will require 225,000 CPU hours.



#### Estimate of Compute, Storage, and Other Resources

We request a total of 2,874,000 ACCESS credits for a one-year allocation from the start date: 2,074,000 core-hours on Expanse, 800,000 core-hours on PSC Bridges-2 RM, 1000 GB of storage on SDSC Expanse, and 1000 GB of storage on PSC Ocean (Bridges-2 storage). This storage has been sufficient in the past, and a breakdown of requested core-hours for each project is detailed in the table below. All core-hours requested involve shared-memory Gaussian 16<sup>6</sup> calculations with up to 64 processors. Reviewers in the

past have questioned the large estimated numbers for some of our projects. The number of calculations for each project is estimated from the number of reaction pathways, the total number of structures for each pathway, including conformers and isomers for each structure, and the necessary calculations in follow-up studies on related systems. The ability to make chemical predictions requires extensive computations to establish the correct reaction pathways and ensure all possibilities are explored sufficiently.

Project	Job type	# of jobs	# of opt <sup>a</sup> rounds	Method	Basis set	# of heavy atoms	# of basis functions <sup>d</sup>	Core-Hours per job <sup>e</sup>	Total Core-Hours <sup>f</sup>
1	Opt <sup>a</sup> + Freq <sup>b</sup>	2000	50	M06-2X	def2-SVP	70	618	160	320,000
	SP <sup>c</sup>	2000	--	M06-2X	def2-QZVPP	70	3165	90	180,000
2	Opt <sup>a</sup> + Freq <sup>b</sup>	1500	50	B3LYP-D3(BJ)	def2-SVP	49	935	240	360,000
	SP <sup>c</sup>	1500	--	B3LYP-D3(BJ)	def2-TZVPP	49	2053	70	105,000
3	Opt <sup>a</sup> + Freq <sup>b</sup>	1000	70	PBE0-D3(BJ)	def2-SVP	139	1491	500	500,000
	SP <sup>c</sup>	1000	--	M06	def2-TZVPP	139	3231	300	300,000
4	Opt <sup>a</sup> + Freq <sup>b</sup>	750	80	$\omega$ B97X-D	def2-SVP	224	2344	600	450,000
	SP <sup>c</sup>	750	--	M06-2X	def2-TZVPP	224	5448	250	188,000
5	Opt <sup>a</sup> + Freq <sup>b</sup>	1200	50	B3LYP-D3(BJ)	def2-SVP	53	350	20	24,000
	SP <sup>c</sup>	1200	--	$\omega$ B97X-D	def2-TZVPP	53	800	18	22,000
6	Opt <sup>a</sup> + Freq <sup>b</sup>	500	55	PBE0-D3(BJ)	6-31*G (SDD for Co)	69	617	180	90,000
	SP <sup>c</sup>	500	--	PBE0-D3(BJ)	6-311**G (SDD for Co)	69	901	220	110,000
7	Opt <sup>a</sup> + Freq <sup>b</sup>	1500	45	PBE0-D3(BJ)	def2-SVP	26	521	105	158,000
	SP <sup>c</sup>	1500	---	PBE0-D3(BJ)	def2-TZVPP	26	1223	45	67,000

<sup>a</sup> Opt = Optimization. <sup>b</sup> Freq = Frequency. <sup>c</sup> SP = Single Point. <sup>d</sup> Calculated from the average number of atoms and the basis set. <sup>e</sup> Estimated from the number of basis functions and core-hours used in benchmark calculations. <sup>f</sup> Rounded up to 1000.

### Software and Specialized Needs

Our research involves *ab initio* density functional theory (DFT) calculations. Most calculations involve rapid conformational searches using CREST<sup>5</sup> (usually costs less than 1000 core-hours per project), followed by geometry optimizations and vibrational frequency calculations at a low theoretical level (e.g. B3LYP<sup>6</sup> with a double- $\zeta$  basis set and D3(BJ) dispersion correction<sup>7, 8</sup>), and finally higher level calculations of single-point energies and solvation energy corrections (e.g.  $\omega$ B97X-D<sup>9</sup> and a triple- $\zeta$  basis set with additional polarization and diffuse functions as well as an SMD<sup>10</sup> or CPCM<sup>11</sup> solvent model). Effective core potential basis sets are used for heavy atoms and transition metals. All of these studies will be performed with shared memory calculations (using up to 64 processors) in Gaussian 16<sup>4</sup> on SDSC Expanse and PSC Bridges-2 RM. No additional support needs are anticipated.

### Team and Team Preparedness

Our team of 14 graduate students, postdocs, undergraduates, and visiting scholars, each of whom requires high-performance supercomputer facilities, has extensive experience with ACCESS and XSEDE. By providing continuous support, ACCESS has been central to our research leading to key accomplishments in many areas of chemistry. Our publications citing ACCESS include innovative developments of computational methods and collaborations with experimentalists throughout the nation and worldwide. The computational resources requested in this proposal are critical to the research performed by our highly experienced team.