

UCLA

UCLA Electronic Theses and Dissertations

Title

Mechanistic Insight from Physical Models of Laboratory-Engineered Catalysts

Permalink

<https://escholarship.org/uc/item/1pd6t43x>

Author

Evans, Declan

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Mechanistic Insight from Physical Models of Laboratory-Engineered Catalysts

A dissertation submitted in partial fulfillment of the requirements
for degree of Doctor of Philosophy in Molecular Biology

by

Declan Marshall Evans

2023

© Copyright by

Declan Marshall Evans

2023

ABSTRACT OF THE DISSERTATION

Mechanistic Insight from Physical Models of Laboratory-Engineered Catalysts

by

Declan Marshall Evans

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2023

Professor Kendall N. Houk, Chair

The development of catalysts for faster, more selective, and more sustainable chemical reactions remains an outstanding goal in chemistry. While advancements in statistical models have greatly improved our predictive ability, physical models remain our most valuable tools for mechanistic understanding. Density functional theory (DFT) calculations provide structural models of reaction intermediates and transition states and can accurately predict thermodynamic quantities. For larger systems, molecular dynamics (MD) simulations using classical force fields model time-dependent movements of protein-ligand complexes. This Thesis combines both approaches to model several designed and engineered catalysts.

In Chapter 2, we investigate dirhodium catalysts that were designed to catalyze an enantioselective Si–H insertion. Our experimental collaborators observed that the enantioselectivity of the reaction correlated strongly with the Hammett constant of the carbene's aryl substituents but lacked a clear understanding as to why. DFT calculations using an $\text{Rh}_2(\text{OAc})_4$ catalyst showed that the aryl rings of the carbene rotate in the transition state, resulting in one aryl conjugating with the vacant carbene *p*-orbital. For asymmetrically substituted carbenes, the transition state with the electron-rich aryl conjugated to the vacant carbene *p*-orbital was always lower in energy, and this

energy difference increased with greater electronic differences between the rings. We then showed that because of this geometric constraint, the chiral environment of the designed catalysts favored one enantiomeric transition state over the other. Again, the energy difference between transition states correlated with the differences in Hammett substituent constants between the rings.

In Chapter 3, we investigate a laboratory-evolved family of flavin-dependent halogenases with orthogonal selectivity for the chlorination of tryptamine. Few mutations separate the wild type from the most distant mutant, and crystal structures show that there are minimal structural differences between the enzymes. DFT calculations and MD simulations established that a catalytic lysine activates HOCl through general acid catalysis for chlorination of the substrate. Energetic differences in the transition states or Wheland intermediates calculated by DFT showed the intrinsic site selectivity for each substrate. Docking calculations and MD simulations of protein-substrate complexes showed how each enzyme binds the substrate to position a single site closest to the halogenating species, influencing the site selectivity of this reaction. Our physical model was then used to predict site selectivity for several nonnative substrates.

Finally, Chapter 4 investigates a set of computationally designed enzymes with varying activity for a chemiluminescence reaction. Initial designs were optimized by site-saturation mutagenesis, and it was not clear why the final proteins were so much more active than the initial designs. Our DFT calculations confirmed the proposed reaction mechanism and showed that anion formation in the substrate was critical for the reaction. Docking calculations and MD simulations showed that LuxSit-i, the most active variant, was the best at stabilizing this anion in the transition states. Changing the substituent resulted in worse stabilization during the simulations, which is consistent with the observed substrate specificity.

All three of these chapters explore notable examples of catalyst engineering and provide physical models for their mechanisms. The insight gained here will hopefully lead to future generations of engineered catalysts with continuously improved capabilities.

The dissertation of Declan Marshall Evans is approved.

Todd O. Yeates

James U. Bowie

Lin Jiang

Philippe Sautet

Kendall N. Houk, Committee Chair

University of California, Los Angeles

2023

Table of Contents

Abstract	ii
List of Figures	viii
List of Tables	x
Acknowledgements	xi
Vita	xii
1 Introduction	1
1.1 Directed Evolution	3
1.2 Computational Design	4
1.3 Density Functional Theory	5
1.4 Molecular Dynamics	7
1.5 Machine Learning	8
References	13
2 Enantioselective Diarylcarbene Insertion into Si–H Bonds Induced by Electronic Properties of the Carbenes	24
2.1 Author List and Affiliation	25
2.2 Abstract	25
2.3 Introduction	26
2.4 Results and Discussion	27
2.5 Mechanistic Studies	29
2.6 Conclusion	31
2.7 Methods	31
2.7.1 General Information	31
2.7.2 Preparation of substrates	32
2.7.3 Preparation of catalysts	34
2.7.4 Typical procedure for Si-H bond insertion reaction	37
2.7.5 DFT Calculations	38
2.8 Computational Models	46
References	83

3	Analysis of Laboratory-Evolved Flavin-Dependent Halogenases Affords a Computational Model for Predicting Halogenase Site Selectivity	89
3.1	Author List and Affiliation	90
3.2	Abstract	90
3.3	Introduction	91
3.4	Results	92
3.4.1	Reversion of Mutations Reveals Residues that Improve Catalysis by Evolved RebH Variants	92
3.4.2	Crystal Structures of the 0S-Tryptamine-FAD, 10S-FAD, and 8F-FAD Complexes Provide Insight into the Effects of Key Mutations on Catalysis	94
3.4.3	Computational Analysis Suggests that Bound HOCl is the Active Halogenating Species in FDH Catalysis	95
3.4.4	DFT Calculations and MD Simulations can be used to Model the Selectivity of RebH Variants	97
3.4.5	Predicting Site Selectivity in Non-native Substrates Beyond Tryptamine	99
3.5	Discussion	101
3.5.1	Crystal Structures, Reversion Mutations, and MD Simulations Provide Molecular Insight into the Altered Selectivity of Evolved FDHs	101
3.5.2	Analysis of Laboratory-Evolved Enzymes Affords a Computational Model for FDH Selectivity	103
3.5.3	The FDH Selectivity Model Accurately Predicts Selectivity on Non-Native Substrates	104
3.6	Conclusion	105
3.7	Methods	107
3.7.1	Materials	107
3.7.2	General Procedures	107
3.7.3	UHPLC/LC-MS Method	108
3.7.4	Cloning RebH Reversion Variants	108
3.7.5	Expression and Purification of MBP-RebF and RebH	109
3.7.6	Specific experimental procedures	109
3.7.7	Crystallization and structure determination	110
3.7.8	Density functional theory calculations	111
3.7.9	pKa calculations	111
3.7.10	Docking calculations	111
3.7.11	Molecular dynamics simulations	111
3.8	Computational Models	141
	References	148
4	De Novo Design of Luciferases Using Deep Learning	155
4.1	Author List and Affiliation	156
4.2	Abstract	156
4.3	Main	157
4.3.1	Family-wide hallucination	158
4.3.2	De novo design of luciferases for DTZ	159
4.3.3	Identification of active luciferases	160

4.3.4	De novo design of luciferases for h-CTZ	161
4.3.5	Luciferase activity optimization	162
4.3.6	Cell imaging and multiplexed bioassay	163
4.4	Conclusion	164
4.5	Methods	166
4.5.1	Materials and general methods	166
4.5.2	General procedures for protein production and purification	167
4.5.3	Computational design of idealized scaffolds	167
4.5.4	RIFdock tuning files	173
4.5.5	Designing theozyme architectures into de novo NTF2 scaffolds	173
4.5.6	Structure prediction of LuxSit with AlphaFold2 and comparison to design model	175
4.5.7	Computational design and characterization of de novo luciferases for h-CTZ	175
4.5.8	Computational SSM to estimate mutation binding free energy	177
4.5.9	Construction and screening of designed luciferase libraries	178
4.5.10	Construction and evaluation of LuxSit site saturation mutagenesis libraries	179
4.5.11	In vitro characterization of photoluminescence properties	180
4.5.12	Circular dichroism (CD)	181
4.5.13	Mammalian cell culture and transfection	181
4.5.14	Fluorescence Microscopy and image analysis	182
4.5.15	Multiplex dual-luciferase reporter assay for the cAMP/PKA and NF- κ B pathways	182
4.5.16	Quantum mechanics calculation for the energy profile of DTZ luminescence	183
4.5.17	Molecular dynamics simulations of proposed enzyme-ligand complexes . .	183
4.5.18	Statistical analysis and reproducibility	185
4.6	Computational Models	207
	References	215
	Conclusion	222
	References	226

List of Figures

1.1	Chapter 2 Abstract	10
1.2	Chapter 3 Abstract	11
1.3	Chapter 4 Abstract	12
2.1	Chiral differentiations of prochiral faces and enantioselective diarylcarbene insertion into Si–H bonds	40
2.2	Enantioselective Si–H Bond Insertion of 4-Nitrophenylphenyl Diazomethane Catalyzed by Chiral Dirhodium Catalysts	41
2.3	Substrate Scope of Diarylcarbene Insertion into Si–H Bonds	42
2.4	Computational studies	43
3.1	Substrate positioning within flavin-dependent halogenases allows for site-selective halogenation	113
3.2	Summary of reversion mutations	114
3.3	Comparison of crystal structures between variants	115
3.4	MD simulations of possible chlorinating species	116
3.5	Transition state structures for the acid-catalyzed chlorination of indole	117
3.6	Substrate binding poses	118
3.7	Comparison of substrate binding in variants	119
S3.1	Molecular dynamics simulations of RebH	125
S3.2	Molecular dynamics simulations of 0S	126
S3.3	Comparison of substrate binding poses in variants with other natural halogenases	127
S3.4	Molecular dynamics simulations of variants 10S and 8F	128
S3.5	Calculated $\Delta\Delta G_{\text{whel}}$ for different sites on the indole ring	129
S3.6	Calculated docking poses for substrate 1	130
S3.7	Calculated docking poses for substrate 2	131
S3.8	Calculated docking poses for substrate 3	132
S3.9	Calculated docking poses for substrate 4	133
S3.10	Calculated docking poses for substrate 5	134
S3.11	Calculated docking poses for substrate 6	135
S3.12	Calculated docking poses for substrate 7	136
S3.13	Full carbon numbering for each nonnative substrate used in this study	137
S3.14	Dynamics of a hydrogen bond triad	138
S3.15	Dynamics of substrate pi stacking	139
S3.16	Comparison of methods	140

4.1	Generation of idealized scaffolds and computational design of de novo luciferases	186
4.2	Biophysical characterization of LuxSit	188
4.3	Characterization of de novo luciferase activity in vitro and in human cells	190
4.4	High substrate specificity of de novo luciferases allows multiplexed bioassay	191
E4.1	Proposed catalytic mechanism of coelenterazine-utilizing luciferases	193
E4.2	Schematic representative of colony-based luciferase screening	194
E4.3	Expression, purification, and structural characterization of LuxSit variants	195
E4.4	Expression, purification, and activity measurement of selected de novo designed luciferases for h-CTZ	197
E4.5	Predicted changes in substrate binding free energy from binding site	198
E4.6	Screening of a randomized NNK library at 60, 96, and 110 positions and sequence alignment between LuxSit and its variants	199
E4.7	Additional characterization of LuxSit variants	200
E4.8	Free energy profile of DTZ chemiluminescence and molecular dynamics simulations of proposed protein-intermediate complexes	201
E4.9	Expression, localization, and luminescence activity of LuxSit-i in live HEK293T and HeLa cells	202
E4.10	Substrate specificity of LuxSit-i and spectrally resolved luciferase-luciferin pairs allow multiplexed bioassay	203
S4.1	Additional computational analysis of designed luciferases	204
S4.2	Additional QM and MM computational models used in this study	205
S4.3	The raw light outputs plot of Figure 4.4c	206

List of Tables

2.1	Rhodium-Catalyzed Enantioselective Si–H Bond Insertions of Diphenyl Diazomethylenes D1–D12	44
2.2	Rhodium-Catalyzed Enantioselective Si–H Bond Insertion of Diphenyl Diazomethanes D13–D30	45
S2.1	Computed Energies of Dirhodium-Tetraformate Transition States	46
S2.2	Computed Dnergies of (S)-C1 Transition States	47
3.1	Kinetic parameters of RebH and variants	120
3.2	Thermodynamic parameters for electrophilic attack of different Cl ⁺ donors on indole	121
3.3	Substrate (ΔG_{TS}^\ddagger) vs catalyst (ΔG_N°) control over site selectivity of tryptamine in RebH and variants	122
3.4	Substrate ($\Delta\Delta G_{\text{whel}}$) vs catalyst ($\Delta\Delta G_N^\circ$) control over site selectivity of non-native substrates in RebH	124
S3.1	Computed Energies of Computational Models	141
S4.1	Computed Energies of Computational Models	207

Acknowledgements

This work was supported in part by the NSF (CHE-1764328, to Ken Houk). Additional support by the NSF under the CCI Center for Selective C–H Functionalization (CCHF, CHE-1205646 to Ken Houk). Additional support by the UCLA Chemistry-Biology Interface training grant (USPHS National Research Service Award T32GM008496) and UCLA Molecular Biology department. Special gratitude to Dr. Audree V. Fowler for her endowment to the Fowler Fellowship in Protein Science. Calculations were performed on the Hoffman2 cluster at UCLA and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant no. ACI-1548562.

I would like to thank Ken and my committee for being incredible mentors and helping me grow into the scientist I am today. I would also like to thank my parents, Karen and Marshall, and my brother, Julian, for their unconditional love and support. My girlfriend, Katie, who has been an incredible partner through this wild journey together. Our cats, Dory and Nutella, for being amazing coworkers throughout the pandemic and always brightening my day. Thank you to all of my friends who have stayed with me over the years and thank you to the new ones I have made along the way. This has been such an incredible experience and I am so grateful to have done it the way that I did, surrounded by the people that I love.

Declan Marshall Evans

LinkedIn • Google Scholar • ResearchGate • ORCID

Research Experience

- 2017- Graduate Student, The University of California, Los Angeles
2014-2017 Undergraduate Researcher, The Pennsylvania State University

Education

- 2017 B.S. in Biochemistry and Molecular Biology, The Pennsylvania State University
2017 B.S. in Mathematics, The Pennsylvania State University

Grants, Honors & Awards

- 2022 Audree V. Fowler Fellowship in Protein Science
2018-2021 UCLA CBI Training Fellowship
2015 Erickson Discovery Grant

Publications

- 2023 A. H. W. Yeh, C. Norn, Y. Kipnis, D. Tischer, S. J. Pellock, D. Evans, P. Ma, G. R. Lee, J. Z. Zhang, I. Anishchenko, B. Coventry, L. Cao, J. Dauparas, S. Halabiya, M. DeWitt, L. Carter, K. N. Houk and D. Baker, *Nature*, accepted
- 2022 M. C. Andorfer[‡], D. Evans[‡], S. Yang, C. Q. He, A. M. Girlich, J. Vergara-Coll, N. Sukumar, K. N. Houk, and J. C. Lewis, *Chem Catalysis*, **2022**, 2(10), 2658-2674
- 2022 I. Benavides, E. D. Raftery, A. G. Bell, D. Evans, W. A. Scott, K. N. Houk, and T. J. Deming, *J. Am. Chem. Soc.* **2022**, 144, 9, 4214–4223

- 2022 K. A. Murray[‡], D. Evans[‡], M. P. Hughes, M. R. Sawaya, C. J. Hu, K. N. Houk, and D. Eisenberg, *ACS Nano*, **2022**, 16, 2, 2154–2163
- 2021 W. A. Scott, E. G. Gharakhanian, A. G. Bell, D. Evans, E. Barun, K. N. Houk, and T. J. Deming, *J. Am. Chem. Soc.* **2021**, 143, 43, 18196–18203
- 2020 L. Yang, D. Evans, B. Xu, W.-T. Li, M.-L. Li, S.-F. Zhu, K. N. Houk, and Q.-L. Zhou, *J. Am. Chem. Soc.* **2020**, 142 (28), 12394-12399
- 2019 Y. Yao, C. An, D. Evans, W. Liu, W. Wang, G. Wei, N. Ding, K. N. Houk, and S.-S. Gao, *J. Am. Chem. Soc.* **2019**, 141 (44), 17517-17521
- 2016 R. Acevedo, D. Evans, K. A. Penrod, S. A. Showalter, *Biophysical Journal* **2016**, 110 (12), 2610-2617

Professional Affiliations

- 2021-2022 Graduate Intern, Invizyne, Monrovia, C.A.
- 2016 Undergraduate Intern, Regeneron Pharmaceuticals, Rensselaer, NY

References

K. N. Houk, University of California, Los Angeles
Molecular Sciences Building 5505B 619 Charles E. Young Drive East
Los Angeles, CA 90095 USA
(310) 206-0515

Last updated: February 23, 2023

Chapter 1

Introduction

Catalysts are critical to faster, more efficient chemical reactions and have a projected market size of over \$35 Billion [1]. The development of cheaper, faster, and more selective catalysts remains an outstanding goal in chemistry. While experimentation has always been and will continue to be the most effective means of catalyst development, the rapid development of computer hardware and software continues to enhance the utility of computation. Advanced statistical models have recently exploded in popularity; some can now make predictions with super-human ability. Despite the utility of this predictive power, statistical models rarely give any mechanistic insight that is critical for understanding by a human. For this insight, we depend on calculated physical models, which provide details about chemical structure and reactivity that are not possible with experimentation alone. In this Thesis, we will discuss the role of physical models in catalyst development and provide examples of catalysts that have been developed through joint experimentation and computation.

Considerable advances have been made in organic and inorganic catalysis, enabling the synthesis of countless pharmaceutical and industrial chemicals [2, 3, 4, 5]. In Chapter 2, a collaboration with Qi-Lin Zhou published in the *Journal of the American Chemical Society*, we investigate a dirhodium catalyst that was developed for enantioselective Si–H insertion with carbenes (Figure 1.1). Dirhodium catalysts of this kind are used extensively in Si–H and C–H activation chemistry [6] due to their ability to stabilize carbene and nitrene intermediates. Despite the power of these catalysts, many are impractical to use at scale because of solvent constraints, incorporation of rare and expensive metals, or difficulties involved in the synthesis. Biocatalysts, catalysts isolated from living organisms, offer a much more scalable and sustainable alternative due to their mild operating conditions and relative ease of synthesis.

Naturally occurring enzymes possess a wealth of chemical functionality, but their utility is limited by the process of natural selection. Only the enzymes that provide a selective advantage for the organism will persist through generations. Engineering new enzymes for non-natural purposes is challenging, as a protein of N residues has 20^N possible combinations of amino acids. Even the smallest peptides have too many possible sequences to observe with exhaustive sampling. Several

enzyme engineering methods have emerged to navigate this enormous search space and the most successful of which will be covered in this Thesis. In Chapter 3, a collaboration with Jared Lewis published in *Chem Catalysis*, we investigate a family of enzymes generated through **Directed evolution** (Figure 1.2). This process involves iterative rounds of mutagenesis and screening and is increasingly used to re-engineer natural enzymes for non-natural purposes. In Chapter 4, a collaboration with David Baker accepted at *Nature*, we investigate an enzyme developed through **Computational Design** (Figure 1.3).

Each study discussed here presents both experimental observations and a computational investigation of the associated mechanism. Quantum Mechanical (QM) models using **Density Functional Theory** (DFT) calculations are necessary to determine transition state structure and explain differences in reactivity. While DFT is used whenever possible, unfavorable scaling with system size necessitates the use of classical molecular mechanics and force fields. While these calculations are unable to observe transition state behavior, they are fast enough to observe how large macromolecules can affect reaction intermediates. Force fields are most used in **Molecular Dynamics** (MD) simulations, which model time-dependent motions of catalyst-substrate complexes. Finally, the explosion in popularity of **Machine learning** methods necessitates their mention here. These models can incorporate large amounts of data to identify patterns in high-dimensional spaces and have led to some of the most important breakthroughs in recent years. Each of these methods will be discussed further, with a particular emphasis on how they relate to catalyst engineering.

1.1 Directed Evolution

The most successful and robust method for enzyme engineering is directed evolution, which imitates the natural process of evolution through iterative rounds of mutation and screening for some desired trait. Almost any trait of the protein can be optimized with this method, including stability [7, 8], solubility [9, 10, 11], activity [12, 13, 14, 15], and selectivity [16, 17]. Researchers at Merck recently demonstrated that directed evolution can be used to design an entire biocatalytic

cascade for the production of pharmaceuticals [18]. Natural enzymes involved in the bacterial purine nucleoside degradation pathway were evolved to accept synthetic precursors and the pathway was run in reverse to synthesize islatravir. The resulting three-step biocatalytic cascade uses fewer than half of the steps as comparable synthetic routes and delivers an impressive 51% overall yield. Moreover, this biocatalytic cascade generates 4 stereocenters with high enantiomeric excess from achiral substrates. The success of this study is largely due to the structural similarity between islatravir and natural purine nucleosides. Directed evolution is generally most successful when the reaction of interest is similar to the natural reaction and there is a high degree of activity already present in the enzyme.

Nonetheless, the evolution of heme-containing enzymes has resulted in enzymes that catalyze reactions not observed in nature [19]. These enzymes represent the potential of what can be achieved through this powerful process. The biggest limitation of directed evolution is the reliance on high-throughput experimentation to generate and screen large libraries of mutants. The incorporation of experiments [20] and machine learning models [21] have been increasingly adopted to alleviate the experimental bottleneck. Despite these limitations, directed evolution will likely remain the dominant method of enzyme engineering for the foreseeable future because it is not limited by the gaps in our structural and mechanistic understanding of enzymes.

1.2 Computational Design

As mentioned earlier, directed evolution works best when an existing enzyme has native activity for the reaction of interest. If no such enzyme exists, or if the expression and isolation of this enzyme is challenging, a protein can be created using computational design methods. These proteins are usually designed following the hypothesis that an enzyme will fold to position catalytic residues for maximum transition state stabilization [22, 23, 24, 25]. Unlike directed evolution, computational design depends on a physical model of the protein and a detailed understanding of the reaction mechanism. Despite advancements in energetic score functions [26, 27] and neu-

ral networks [28, 29, 30, 31], the gaps in our understanding still result in low success rates for computationally designed enzymes.

The most successful examples of computational enzyme design usually rely on additional experimental methods to improve activity. For example, the Kemp Eliminase originally designed by Baker and Houk [24] underwent multiple rounds of directed evolution to afford a highly active enzyme having little in common with the initial design [12, 13, 14, 15]. Even the luciferase described in detail in Chapter 4 used site saturation mutagenesis to achieve the most active mutant. Despite these limitations, computational design will only improve as neural networks and physical models improve.

1.3 Density Functional Theory

Nearly a century ago, physicist Paul A. M. Dirac famously claimed that “the underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.” [32] This bold claim postulated that our physical understanding of atomic structure was sufficient to answer any chemical question, the only bottleneck was the computational power. No doubt computational power has increased substantially in the almost 100 years since Dirac’s claim, yet the exact application of quantum mechanics (QM) remains possible in only the smallest systems. This is primarily due to the difficulties associated with solving the many-bodied Schrödinger equation.

The advent of density functional theory [33, 34] addressed this problem directly by operating on the electron density rather than the wavefunction. While a universal density functional is known to exist, no such functional has been determined to date. Instead, approximate density functionals have been widely developed to accurately model certain types of systems. Extensive benchmarking

studies have been published that compare the performance of different functionals on different systems [35, 36, 37]. By far the most used density functional in biological and organic chemistry is B3LYP [38] due to its favorable cost-accuracy trade-off. Other methods like M06-2X [39] and ω B97X-D [40] are also used regularly in biological systems, whereas M06 [39] and PBE0 [41, 42] are commonly used for organometallics. In recent years, the accuracy of B3LYP and many other functionals have been improved by the incorporation of empirical corrections to dispersion, leading to the so-called DFT-D3 methods [43].

This report will focus primarily on calculations using Gaussian [44, 45], however, several other computational chemistry software packages like Orca [46, 47] and Q-Chem [48] are commonly used as well. Gaussian calculations typically start with a geometry optimization, where an input set of atomic coordinates are adjusted following the Berny optimization algorithm until a stationary point is reached on the potential energy surface [49]. A subsequent frequency calculation, which calculates the second derivatives of the energy with respect to the Cartesian nuclear coordinates, determines whether this stationary point represents a ground state (local minimum) or transition state (first-order saddle point). This information can then be used along with quasi-harmonic corrections [50, 51, 52] to obtain thermodynamic values like free energy.

Despite the benefits of DFT, most standard code implementations scale exponentially with the size of the system. This poor scaling makes DFT impractical for systems larger than a few hundred atoms. Semi-empirical quantum mechanical (SQM) methods, like GFN2-xTB [53], are popular for systems of this size due to their increased calculation speed. These methods work similarly to DFT, but several of the most computationally expensive steps are replaced with data obtained from experiments. The speed of these SQM calculations also makes them attractive for conformational searches like CREST [54], which uses GFN2-xTB and a meta-dynamics driven search algorithm to rapidly explore conformational space for the global minimum. Despite the efficiency of these methods, they are still impractical for very large macromolecular complexes like proteins. In such systems, we must turn to other physical models which scale linearly with system size.

1.4 Molecular Dynamics

Modern computational resources are not yet capable of modeling proteins with QM methods on a practical timescale. These systems require methods that utilize classical mechanics, also called molecular mechanics (MM), which scale linearly with system size. Energy is calculated from a linear combination of pre-parametrized energy functions (force fields) which are obtained from experiments or higher accuracy QM methods. While some force fields are parametrized for transition states [55], the majority are parametrized only for ground states, meaning bonds cannot break or form using classical molecular mechanics. Furthermore, the way these force fields are parameterized is variable, and different force fields have been developed to model specific systems. Several biological force fields have been developed, including CHARMM [56] and AMOEBA [57, 58, 59, 60, 61]. The Rosetta force field [26], which includes both physical and statistical parameters, has also gained popularity in recent years for protein design and engineering. This report will focus primarily on applications of the AMBER force field [62, 63, 64, 27] which is parametrized for proteins and widely implemented in most MD software. The general AMBER force field (GAFF) [65] is an extension of AMBER which uses charges derived from DFT calculations to parametrize force fields for any ligands or cofactors.

Molecular dynamics simulations begin with a structure of some protein-ligand complex. If no such complex is available from experimental structural characterization, the ligand can be docked into the protein computationally. Several docking programs exist, the most popular being AutoDock Vina [66] or Rosetta [67, 68, 69]. These programs use low-resolution force fields to rapidly sample different conformations of the protein-ligand complex and predict the most likely binding orientation.

Starting from a protein-ligand complex, the system is solvated and minimized using the steepest descent or conjugate gradient methods. Forces are randomly assigned to each atom based on the total energy of the system, and each atom accelerates due to its corresponding force and mass. Forces are recalculated from the force field at discrete timesteps throughout the simulation (usu-

ally 1 or 2 fs) and the acceleration of each atom is adjusted accordingly. Recent advancements in graphics processing units (GPUs) and hardware specialized for MD [70, 71, 72] have dramatically improved calculation speeds, allowing simulation times on the nanosecond to microsecond timescales for most proteins and macromolecules. While these simulations never show chemical reactivity, they can give critical insight into how the enzyme can bind to and influence reaction intermediates. In some cases, it is beneficial to use both QM and MM simultaneously in so-called QM/MM. These methods treat part of the system, usually the active site, with quantum mechanics and the rest of the system with force fields. This report will not cover these methods, with the exception of ONIOM [73], which is used in Gaussian to conduct multi-layered QM/QM or QM/MM Gaussian calculations.

1.5 Machine Learning

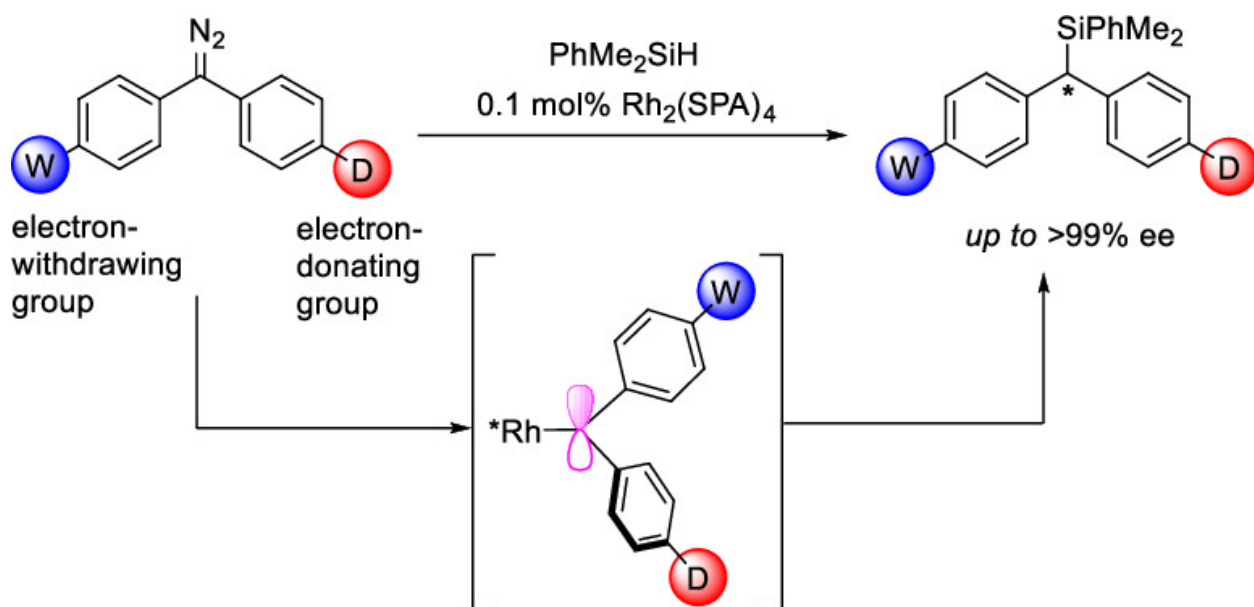
Machine learning has revolutionized nearly every scientific field in the last decade and computational chemistry is no exception. As stated earlier, these models can incorporate large amounts of data to identify patterns in high-dimensional spaces with super-human ability. Perhaps the most impressive examples are the neural networks developed to predict protein structures as accurately as experimental methods [28, 29, 30, 31]. These networks identify patterns in protein sequences and predict structural constraints to guide high-resolution force field minimizations. Accurate structures of proteins [74] and protein complexes [75] are now widely available. Neural networks are also being increasingly used in protein design [76, 77, 78, 79, 80] and will be discussed in Chapter 4.

There are countless examples of models that read experimental results or DFT-calculated values to estimate transition state barrier heights [81, 82, 83, 84, 85, 86, 87, 88]. These models can sometimes provide barrier height predictions on par with DFT yet use only a fraction of the compute time. Similar works targeted at reducing DFT compute time have developed semi-empirical quantum mechanics methods with DFT-level accuracy [89], new DFT functionals [90], and even

entirely machine-learned potentials [91]. Machine learning has also been used in synthesis to predict the yields of reactions [92] and even optimize reaction conditions [93]. While these methods rarely improve our mechanistic understanding, their predictive ability makes them invaluable tools.

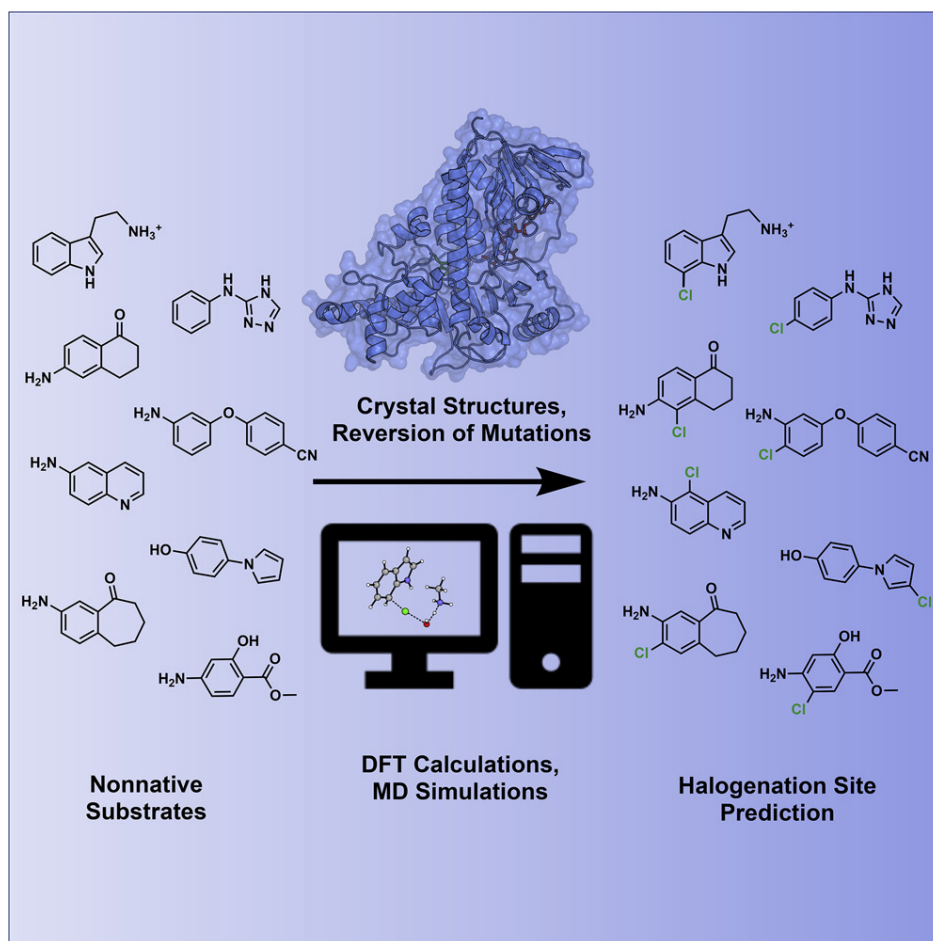
The next chapters each give an example of catalysts that were developed with the above methods and the insight we gained from computational investigation. In Chapter 2 we investigate an organometallic catalyst developed through traditional chemical synthesis to catalyze an enantioselective Si-H insertion reaction. Computational models using QM and QM/MM methods were used and show that the catalyst provides a chiral environment that restricts the conformations of the substrate during the reaction. In Chapter 3 we investigate a family of enzymes that were evolved through directed evolution to have orthogonal selectivity for chlorination of low molecular weight arenes. Small model theozyme calculations using QM showed intrinsic differences in substrate reactivity, and MD simulations with protein-ligand complexes showed how the enzyme alters this intrinsic reactivity. In Chapter 4 we investigate a set of computationally designed enzymes that catalyze the chemiluminescence of diphenylcoelenterazine, a synthetic marine luciferase. Small model theozyme calculations using QM confirmed the proposed reaction mechanism, and MD simulations of protein-ligand complexes showed how mutations in the active site affected intermediate binding in the most active variants. Finally, we end with a summary of each project and an outlook on the future of catalyst development.

Figure 1.1: Chapter 2 Abstract



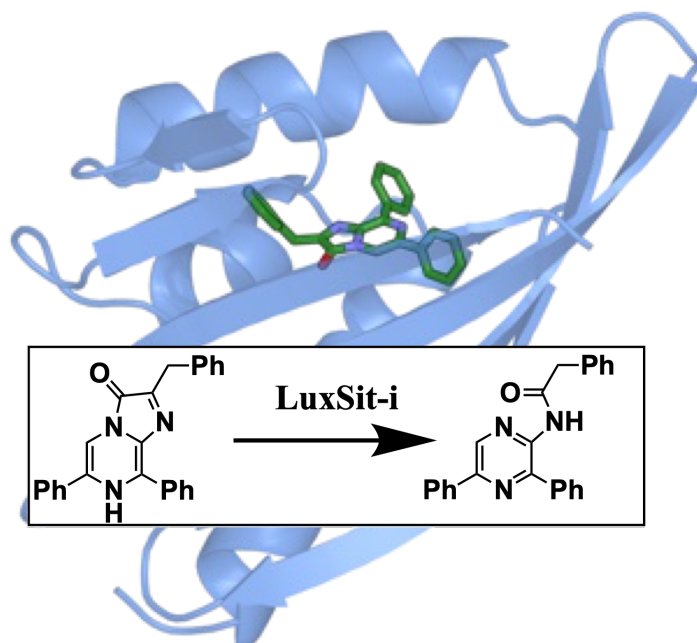
Catalytic enantioselection usually depends on differences in steric interactions between prochiral substrates and a chiral catalyst. We have discovered a carbene Si–H insertion in which the enantioselectivity depends primarily on the electronic characteristics of the carbene substrate, and the $\log(er)$ values are linearly related to Hammett parameters. A new class of chiral tetraphosphate dirhodium catalysts was developed; it shows excellent activity and enantioselectivity for the insertion of diarylcarbenes into the Si–H bond of silanes. Computational and mechanistic studies show how the electronic differences between the two aryls of the carbene lead to differences in energies of the diastereomeric transition states. This study provides a new strategy for asymmetric catalysis exploiting the electronic properties of the substrates.

Figure 1.2: Chapter 3 Abstract



Flavin-dependent halogenases (FDHs) catalyze selective halogenation of electron-rich aromatic compounds without the need for harsh oxidants required by conventional oxidative halogenation reactions. Predictive models for halogenase site selectivity could greatly improve their utility for chemical synthesis. Toward this end, we analyzed the structures and selectivity of three halogenase variants evolved to halogenate tryptamine with orthogonal selectivity. Crystal structures and reversion mutations revealed key residues involved in altering halogenase selectivity. Density functional theory calculations and molecular dynamics simulations are both consistent with hypohalous acid as the active halogenating species in FDH catalysis. This model was used to accurately predict the site selectivity of halogenase variants toward different synthetic substrates, providing a valuable tool for implementing halogenases in biocatalysis efforts.

Figure 1.3: Chapter 4 Abstract



De novo enzyme design has sought to introduce active sites and substrate binding pockets predicted to catalyze a reaction of interest into geometrically compatible native scaffolds [94, 95], but has been limited by a lack of suitable protein structures and the complexity of native protein sequence-structure relationships. Here we describe a deep-learning based “family-wide hallucination” approach that generates large numbers of idealized protein structures containing diverse pocket shapes and designed sequences that encode them. We use these scaffolds to design artificial luciferases that selectively catalyze the oxidative chemiluminescence of the synthetic luciferin substrates diphenylterazine (DTZ) [96] and 2-deoxycoelenterazine (h-CTZ) through the placement of an arginine guanidinium group adjacent to an anion species that develops during the reaction in a high shape complementarity binding pocket. For both luciferin substrates, we obtain designed luciferases with high selectivity; the most active of these is a small (13.9 kDa) and thermostable ($T_M > 95^\circ\text{C}$) enzyme with a catalytic efficiency on DTZ ($k_{\text{cat}}/K_M = 106 \text{ M}^{-1}\text{s}^{-1}$) comparable to native luciferases but with much higher substrate specificity. The design of highly active and specific biocatalysts from scratch with broad applications in biomedicine is an important milestone for computational enzyme design, and our approach should enable the design of a wide range of novel luciferases and other enzymes.

References

- [1] Catalyst market size, share and trends analysis report by raw material (chemical compounds, zeolites, metals), by product (heterogeneous, homogeneous), by application, by region, and segment forecasts, 2020 - 2027 (978-1-68038-228-0). *Grand View Research*.
- [2] Ziwei Song, Wei Jin, Faming Gao, and Xin Jin. Recent advances in catalyst development for transesterification of dialkyl carbonates with phenol. *Industrial & Engineering Chemistry Research*, 59(47):20630–20645, 2020.
- [3] Huazhang Liu. Ammonia synthesis catalyst 100 years: Practice, enlightenment and challenge. *Chinese Journal of Catalysis*, 35(10):1619–1640, 2014.
- [4] Jens Sehested. Industrial and scientific directions of methanol catalyst development. *Journal of Catalysis*, 371:368–375, 2019.
- [5] Hessam Jahangiri, James Bennett, Parvin Mahjoubi, Karen Wilson, and Sai Gu. A review of advanced catalyst development for fischer–tropsch synthesis of hydrocarbons from biomass derived syn-gas. *Catalysis Science & Technology*, 4(8):2210–2229, 2014.
- [6] Huw ML Davies and Kuangbiao Liao. Dirhodium tetracarboxylates as catalysts for selective intermolecular c–h functionalization. *Nature Reviews Chemistry*, 3(6):347–360, 2019.
- [7] Chang Ren, Xin Wen, Jun Mencius, and Shu Quan. Selection and screening strategies in directed evolution to improve protein stability. *Bioresources and Bioprocessing*, 6(1):1–14, 2019.
- [8] Elizabeth L Bell, Ross Smithson, Siobhan Kilbride, Jake Foster, Florence J Hardy, Saranarayanan Ramachandran, Aleksander A Tedstone, Sarah J Haigh, Arthur A Garforth, Philip JR Day, et al. Directed evolution of an efficient and thermostable pet depolymerase. *Nature Catalysis*, 5(8):673–681, 2022.
- [9] Susanne Van Den Berg, Per-Åke Löfdahl, Torleif Härd, and Helena Berglund. Improved solubility of tev protease by directed evolution. *Journal of biotechnology*, 121(3):291–298, 2006.
- [10] Jian-Wei Liu, Kieran S Hadler, Gerhard Schenk, and David Ollis. Using directed evolution to improve the solubility of the c-terminal domain of escherichia coli aminopeptidase p: Implications for metal binding and protein stability. *The FEBS Journal*, 274(18):4742–4751, 2007.
- [11] Tina Wang, Ahmed H Badran, Tony P Huang, and David R Liu. Continuous directed evolution of proteins with improved soluble expression. *Nature chemical biology*, 14(10):972–980, 2018.
- [12] Olga Khersonsky, Daniela Röthlisberger, Orly Dym, Shira Albeck, Colin J. Jackson, David Baker, and Dan S. Tawfik. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the ke07 series. *Journal of Molecular Biology*, 396(4):1025–1042, 2010.

- [13] Olga Khersonsky, Daniela Röthlisberger, Andrew M. Wollacott, Paul Murphy, Orly Dym, Shira Albeck, Gert Kiss, K.N. Houk, David Baker, and Dan S. Tawfik. Optimization of the in-silico-designed kemp eliminase ke70 by computational design and directed evolution. *Journal of Molecular Biology*, 407(3):391–412, 2011.
- [14] Olga Khersonsky, Gert Kiss, Daniela Röthlisberger, Orly Dym, Shira Albeck, Kendall N. Houk, David Baker, and Dan S. Tawfik. Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed kemp eliminase ke59. *Proceedings of the National Academy of Sciences*, 109(26):10358–10363, 2012.
- [15] Renee Otten, Ricardo A. P. Pádua, H. Adrian Bunzel, Vy Nguyen, Warintra Pitsawong, MacKenzie Patterson, Shuo Sui, Sarah L. Perry, Aina E. Cohen, Donald Hilvert, and Dorothee Kern. How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science*, 370(6523):1442–1446, 2020.
- [16] Mary C. Andorfer, Hyun June Park, Jaylie Vergara-Coll, and Jared C. Lewis. Directed evolution of rebh for catalyst-controlled halogenation of indole c–h bonds. *Chem. Sci.*, 7:3720–3729, 2016.
- [17] Xiaoqiang Jia, Yubing Ma, Rongrong Bu, Tingting Zhao, and Kang Wu. Directed evolution of a transcription factor pbrr to improve lead selectivity and reduce zinc interference through dual selection. *Amb Express*, 10(1):1–10, 2020.
- [18] Mark A. Huffman, Anna Fryszkowska, Oscar Alvizo, Margie Borra-Garske, Kevin R. Campos, Keith A. Canada, Paul N. Devine, Da Duan, Jacob H. Forstater, Shane T. Grosser, Holst M. Halsey, Gregory J. Hughes, Junyong Jo, Leo A. Joyce, Joshua N. Kolev, Jack Liang, Kevin M. Maloney, Benjamin F. Mann, Nicholas M. Marshall, Mark McLaughlin, Jeffrey C. Moore, Grant S. Murphy, Christopher C. Nawrat, Jovana Nazor, Scott Novick, Niki R. Patel, Agustina Rodriguez-Granillo, Sandra A. Robaire, Edward C. Sherer, Matthew D. Truppo, Aaron M. Whittaker, Deeptak Verma, Li Xiao, Yingju Xu, and Hao Yang. Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science*, 366(6470):1255–1259, 2019.
- [19] Yang Yang and Frances H. Arnold. Navigating the unnatural reaction space: Directed evolution of heme proteins for selective carbene and nitrene transfer. *Accounts of Chemical Research*, 54(5):1209–1225, 2021. PMID: 33491448.
- [20] Sagar Bhattacharya, Eleonora G Margheritis, Katsuya Takahashi, Alona Kulesha, Areetha D’Souza, Inhye Kim, Jennifer H Yoon, Jeremy RH Tame, Alexander N Volkov, Olga V Makhlynets, et al. Nmr-guided directed evolution. *Nature*, 610(7931):389–393, 2022.
- [21] Kevin K Yang, Zachary Wu, and Frances H Arnold. Machine-learning-guided directed evolution for protein engineering. *Nature methods*, 16(8):687–694, 2019.
- [22] Daniel N. Bolon and Stephen L. Mayo. Enzyme-like proteins by computational design. *Proceedings of the National Academy of Sciences*, 98(25):14274–14279, 2001.

- [23] Lin Jiang, Eric A. Althoff, Fernando R. Clemente, Lindsey Doyle, Daniela Röthlisberger, Alexandre Zanghellini, Jasmine L. Gallaher, Jamie L. Betker, Fujie Tanaka, Carlos F. Barbas, Donald Hilvert, Kendall N. Houk, Barry L. Stoddard, and David Baker. De novo computational design of retro-aldol enzymes. *Science*, 319(5868):1387–1391, 2008.
- [24] Daniela Röthlisberger, Olga Khersonsky, Andrew M. Wollacott, Lin Jiang, Jason DeChancie, Jamie Betker, Jasmine L. Gallaher, Eric A. Althoff, Alexandre Zanghellini, Orly Dym, Shira Albeck, Kendall N. Houk, Dan S. Tawfik, and David Baker. Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192):190–195, May 2008.
- [25] Justin B. Siegel, Alexandre Zanghellini, Helena M. Lovick, Gert Kiss, Abigail R. Lambert, Jennifer L. St.Clair, Jasmine L. Gallaher, Donald Hilvert, Michael H. Gelb, Barry L. Stoddard, Kendall N. Houk, Forrest E. Michael, and David Baker. Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. *Science*, 329(5989):309–313, 2010.
- [26] Rebecca F Alford, Andrew Leaver-Fay, Jeliasko R Jeliaskov, Matthew J O’Meara, Frank P DiMaio, Hahnbeom Park, Maxim V Shapovalov, P Douglas Renfrew, Vikram K Mulligan, Kalli Kappel, et al. The rosetta all-atom energy function for macromolecular modeling and design. *Journal of chemical theory and computation*, 13(6):3031–3048, 2017.
- [27] Chuan Tian, Koushik Kasavajhala, Kellon A. A. Belfon, Lauren Raguette, He Huang, Angela N. Migués, John Bickel, Yuzhang Wang, Jorge Pincay, Qin Wu, and Carlos Simmerling. ff19sb: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of Chemical Theory and Computation*, 16(1):528–552, 2020. PMID: 31714766.
- [28] Mohammed AlQuraishi. Alphafold at casp13. *Bioinformatics*, 35(22):4862–4865, 2019.
- [29] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, Aug 2021.
- [30] Jianyi Yang, Ivan Anishchenko, Hahnbeom Park, Zhenling Peng, Sergey Ovchinnikov, and David Baker. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117(3):1496–1503, 2020.
- [31] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alberdina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christopher Garcia, Nick V. Grishin, Paul D. Adams,

- Randy J. Read, and David Baker. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.
- [32] Paul Adrien Maurice Dirac and Ralph Howard Fowler. Quantum mechanics of many-electron systems. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 123(792):714–733, 1929.
- [33] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.
- [34] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.
- [35] Lori A. Burns, Álvaro Vázquez-Mayagoitia, Bobby G. Sumpter, and C. David Sherrill. Density-functional approaches to noncovalent interactions: A comparison of dispersion corrections (dft-d), exchange-hole dipole moment (xdm) theory, and specialized functionals. *The Journal of Chemical Physics*, 134(8):084107, 2011.
- [36] Lars Goerigk, Andreas Hansen, Christoph Bauer, Stephan Ehrlich, Asim Najibi, and Stefan Grimme. A look at the density functional theory zoo with the advanced gmtkn55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.*, 19:32184–32215, 2017.
- [37] Narbe Mardirossian and Martin Head-Gordon. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular Physics*, 115(19):2315–2372, 2017.
- [38] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993.
- [39] Yan Zhao and Donald G. Truhlar. 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. *Theoretical Chemistry Accounts*, 120(1):215–241, May 2008.
- [40] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10:6615–6620, 2008.
- [41] Matthias Ernzerhof and Gustavo E Scuseria. Assessment of the perdedew–burke–ernzerhof exchange–correlation functional. *The Journal of chemical physics*, 110(11):5029–5036, 1999.
- [42] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The pbe0 model. *The Journal of chemical physics*, 110(13):6158–6170, 1999.
- [43] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of Chemical Physics*, 132(15):154104, 2010.

- [44] H. B. Schlegel G. E. Scuseria M. A. Robb J. R. Cheeseman G. Scalmani V. Barone G. A. Petersson H. Nakatsuji X. Li M. Caricato A. Marenich J. Bloino B. G. Janesko R. Gomperts B. Mennucci H. P. Hratchian J. V. Ortiz A. F. Izmaylov J. L. Sonnenberg D. Williams-Young F. Ding F. Lipparini F. Egidi J. Goings B. Peng A. Petrone T. Henderson D. Ranasinghe V. G. Zakrzewski J. Gao N. Rega G. Zheng W. Liang M. Hada M. Ehara K. Toyota R. Fukuda J. Hasegawa M. Ishida T. Nakajima Y. Honda O. Kitao H. Nakai T. Vreven K. Throssell 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 J. M. Millam M. Klene C. Adamo R. Cammi J. W. Ochterski R. L. Martin K. Morokuma O. Farkas J. B. Foresman M. J. Frisch, G. W. Trucks and D. J. Fox. Gaussian 09 Revision D.01, 2016. Gaussian Inc. Wallingford CT.
- [45] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian 16 Revision A.03, 2016. Gaussian Inc. Wallingford CT.
- [46] Frank Neese. The orca program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(1):73–78, 2012.
- [47] Frank Neese. Software update: the orca program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1327, 2018.
- [48] Yihan Shao, Zhengting Gan, Evgeny Epifanovsky, Andrew T.B. Gilbert, Michael Wormit, Joerg Kussmann, Adrian W. Lange, Andrew Behn, Jia Deng, Xintian Feng, Debashree Ghosh, Matthew Goldey, Paul R. Horn, Leif D. Jacobson, Ilya Kaliman, Rustam Z. Khaliullin, Tomasz Kuś, Arie Landau, Jie Liu, Emil I. Proynov, Young Min Rhee, Ryan M. Richard, Mary A. Rohrdanz, Ryan P. Steele, Eric J. Sundstrom, H. Lee Woodcock III, Paul M. Zimmerman, Dmitry Zuev, Ben Albrecht, Ethan Alguire, Brian Austin, Gregory J. O. Beran, Yves A. Bernard, Eric Berquist, Kai Brandhorst, Ksenia B. Bravaya, Shawn T. Brown, David Casanova, Chun-Min Chang, Yunqing Chen, Siu Hung Chien, Kristina D. Closser, Deborah L. Crittenden, Michael Diedenhofen, Robert A. DiStasio Jr., Hainam Do, Anthony D. Dutoi, Richard G. Edgar, Shervin Fatehi, Laszlo Fusti-Molnar, An Ghysels, Anna Golubeva-Zadorozhnaya, Joseph Gomes, Magnus W.D. Hanson-Heine, Philipp H.P. Harbach, Andreas W. Hauser, Edward G. Hohenstein, Zachary C. Holden, Thomas-C. Jagau, Hyunjun Ji, Benjamin Kaduk, Kirill Khistyayev, Jaehoon Kim, Jihan Kim, Rollin A. King, Phil Klunzinger, Dmytro Kosenkov, Tim Kowalczyk, Caroline M. Krauter, Ka Un Lao, Adèle D. Laurent, Keith V. Lawler, Sergey V. Levchenko, Ching Yeh Lin, Fenglai Liu, Ester Livshits, Ro-

- hini C. Lochan, Arne Luenser, Prashant Manohar, Samuel F. Manzer, Shan-Ping Mao, Narbe Mardirossian, Aleksandr V. Marenich, Simon A. Maurer, Nicholas J. Mayhall, Eric Neuscamman, C. Melania Oana, Roberto Olivares-Amaya, Darragh P. O'Neill, John A. Parkhill, Trilisa M. Perrine, Roberto Peverati, Alexander Prociuk, Dirk R. Rehn, Edina Rosta, Nicholas J. Russ, Shaama M. Sharada, Sandeep Sharma, David W. Small, Alexander Sodt, Tamar Stein, David Stück, Yu-Chuan Su, Alex J.W. Thom, Takashi Tsuchimochi, Vitalii Vanovschi, Leslie Vogt, Oleg Vydrov, Tao Wang, Mark A. Watson, Jan Wenzel, Alec White, Christopher F. Williams, Jun Yang, Sina Yeganeh, Shane R. Yost, Zhi-Qiang You, Igor Ying Zhang, Xing Zhang, Yan Zhao, Bernard R. Brooks, Garnet K.L. Chan, Daniel M. Chipman, Christopher J. Cramer, William A. Goddard III, Mark S. Gordon, Warren J. Hehre, Andreas Klamt, Henry F. Schaefer III, Michael W. Schmidt, C. David Sherrill, Donald G. Truhlar, Arieh Warshel, Xin Xu, Alán Aspuru-Guzik, Roi Baer, Alexis T. Bell, Nicholas A. Besley, Jeng-Da Chai, Andreas Dreuw, Barry D. Dunietz, Thomas R. Furlani, Steven R. Gwaltney, Chao-Ping Hsu, Yousung Jung, Jing Kong, Daniel S. Lambrecht, WanZhen Liang, Christian Ochsenfeld, Vitaly A. Rassolov, Lyudmila V. Slipchenko, Joseph E. Subotnik, Troy Van Voorhis, John M. Herbert, Anna I. Krylov, Peter M.W. Gill, and Martin Head-Gordon. Advances in molecular quantum chemistry contained in the q-chem 4 program package. *Molecular Physics*, 113(2):184–215, 2015.
- [49] Xiaosong Li and Michael J. Frisch. Energy-represented direct inversion in the iterative subspace within a hybrid geometry optimization method. *Journal of Chemical Theory and Computation*, 2(3):835–839, 2006. PMID: 26626690.
- [50] G. Luchini, Alegre-Requena J. V., Y. Guan, I. Funes-Ardoiz, and R. S. Paton. Goodvibes 3.0.1. 2019.
- [51] Stefan Grimme. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chemistry—A European Journal*, 18(32):9955–9964, 2012.
- [52] Yi-Pei Li, Joseph Gomes, Shaama Mallikarjun Sharada, Alexis T Bell, and Martin Head-Gordon. Improved force-field parameters for qm/mm simulations of the energies of adsorption for molecules in zeolites and a free rotor correction to the rigid rotor harmonic oscillator model for adsorption enthalpies. *The Journal of Physical Chemistry C*, 119(4):1840–1850, 2015.
- [53] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of Chemical Theory and Computation*, 15(3):1652–1671, 2019. PMID: 30741547.
- [54] Philipp Pracht, Fabian Bohle, and Stefan Grimme. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.*, 22:7169–7192, 2020.
- [55] Adri CT Van Duin, Siddharth Dasgupta, Francois Lorant, and William A Goddard. Reaxff: a reactive force field for hydrocarbons. *The Journal of Physical Chemistry A*, 105(41):9396–9409, 2001.

- [56] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell Jr. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4):671–690, 2010.
- [57] Pengyu Ren and Jay W Ponder. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *Journal of computational chemistry*, 23(16):1497–1506, 2002.
- [58] Pengyu Ren and Jay W Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B*, 107(24):5933–5947, 2003.
- [59] Alan Grossfield, Pengyu Ren, and Jay W Ponder. Ion solvation thermodynamics from simulation with a polarizable force field. *Journal of the American Chemical Society*, 125(50):15671–15682, 2003.
- [60] Pengyu Ren and Jay W Ponder. Temperature and pressure dependence of the amoeba water model. *The Journal of Physical Chemistry B*, 108(35):13427–13437, 2004.
- [61] Pengyu Ren, Chuanjie Wu, and Jay W Ponder. Polarizable atomic multipole-based molecular mechanics for organic molecules. *Journal of chemical theory and computation*, 7(10):3143–3161, 2011.
- [62] D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, and P. A. Kollman. Amber 2016, 2016. University of California, San Francisco.
- [63] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, J. T. Berryman, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, G. A. Cisneros, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, M. C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O’Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J Wang, J. Wang, H. Wei, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York, S. Zhao, , and P. A. Kollman. Amber 2022, 2022. University of California, San Francisco.
- [64] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015. PMID: 26574453.

- [65] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. "development and testing of a general amber force field". *J Comput Chem*, 25(9):1157–1174, jul 2004.
- [66] Oleg Trott and Arthur 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(2):455–461, 2010.
- [67] Jens Meiler and David Baker. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins*, 65(3):538–548, November 2006.
- [68] Ian W Davis and David Baker. RosettaLigand docking with full ligand and receptor flexibility. *J Mol Biol*, 385(2):381–392, November 2008.
- [69] Ian W Davis, Kaushik Raha, Martha S Head, and David Baker. Blind docking of pharmaceutically relevant compounds using RosettaLigand. *Protein Sci*, 18(9):1998–2002, September 2009.
- [70] David E. Shaw, Martin M. Deneroff, Ron O. Dror, Jeffrey S. Kuskin, Richard H. Larson, John K. Salmon, Cliff Young, Brannon Batson, Kevin J. Bowers, Jack C. Chao, Michael P. Eastwood, Joseph Gagliardo, J. P. Grossman, C. Richard Ho, Douglas J. Ierardi, István Kolossváry, John L. Klepeis, Timothy Layman, Christine McLeavey, Mark A. Moraes, Rolf Mueller, Edward C. Priest, Yibing Shan, Jochen Spengler, Michael Theobald, Brian Towles, and Stanley C. Wang. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM*, 51(7):91–97, jul 2008.
- [71] David E. Shaw, J.P. Grossman, Joseph A. Bank, Brannon Batson, J. Adam Butts, Jack C. Chao, Martin M. Deneroff, Ron O. Dror, Amos Even, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Brian Greskamp, C. Richard Ho, Douglas J. Ierardi, Lev Iserovich, Jeffrey S. Kuskin, Richard H. Larson, Timothy Layman, Li-Siang Lee, Adam K. Lerer, Chester Li, Daniel Killebrew, Kenneth M. Mackenzie, Shark Yeuk-Hai Mok, Mark A. Moraes, Rolf Mueller, Lawrence J. Nociolo, Jon L. Peticolas, Terry Quan, Daniel Ramot, John K. Salmon, Daniele P. Scarpazza, U. Ben Schafer, Naseer Siddique, Christopher W. Snyder, Jochen Spengler, Ping Tak Peter Tang, Michael Theobald, Horia Toma, Brian Towles, Benjamin Vitale, Stanley C. Wang, and Cliff Young. Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 41–53, 2014.
- [72] David E. Shaw, Peter J. Adams, Asaph Azaria, Joseph A. Bank, Brannon Batson, Alistair Bell, Michael Bergdorf, Jhanvi Bhatt, J. Adam Butts, Timothy Correia, Robert M. Dirks, Ron O. Dror, Michael P. Eastwood, Bruce Edwards, Amos Even, Peter Feldmann, Michael Fenn, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Maria Grolatova, Brian Greskamp, J.P. Grossman, Justin Gullingsrud, Anissa Harper, William Hasenplaugh, Mark Heily, Benjamin Colin Heshmat, Jeremy Hunt, Douglas J. Ierardi, Lev Iserovich, Bryan L. Jackson, Nick P. Johnson, Mollie M. Kirk, John L. Klepeis, Jeffrey S. Kuskin, Kenneth M. Mackenzie, Roy J. Mader, Richard McGowen, Adam McLaughlin,

- Mark A. Moraes, Mohamed H. Nasr, Lawrence J. Nociolo, Lief O'Donnell, Andrew Parker, Jon L. Peticolas, Goran Pocina, Cristian Predescu, Terry Quan, John K. Salmon, Carl Schwink, Keun Sup Shim, Naseer Siddique, Jochen Spengler, Tamas Szalay, Raymond Tabladillo, Reinhard Tartler, Andrew G. Taube, Michael Theobald, Brian Towles, William Vick, Stanley C. Wang, Michael Wazlowski, Madeleine J. Weingarten, John M. Williams, and Kevin A. Yuh. Anton 3: Twenty microseconds of molecular dynamics simulation before lunch. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [73] Stefan Dapprich, István Komáromi, K.Suzie Byun, Keiji Morokuma, and Michael J Frisch. A new oniom implementation in gaussian98. part i. the calculation of energies, gradients, vibrational frequencies and electric field derivatives. dedicated to professor keiji morokuma in celebration of his 65th birthday.1. *Journal of Molecular Structure: THEOCHEM*, 461-462:1–21, 1999.
- [74] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Židek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, and Sameer Velankar. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1):D439–D444, 11 2021.
- [75] Ian R Humphreys, Jimin Pei, Minkyung Baek, Aditya Krishnakumar, Ivan Anishchenko, Sergey Ovchinnikov, Jing Zhang, Travis J Ness, Sudeep Banjade, Saket R Bagde, et al. Computed structures of core eukaryotic protein complexes. *Science*, 374(6573):eabm4805, 2021.
- [76] Ivan Anishchenko, Samuel J. Pellock, Tamuka M. Chidyausiku, Theresa A. Ramelot, Sergey Ovchinnikov, Jingzhou Hao, Khushboo Bafna, Christoffer Norn, Alex Kang, Asim K. Bera, Frank DiMaio, Lauren Carter, Cameron M. Chow, Gaetano T. Montelione, and David Baker. De novo protein design by deep network hallucination. *Nature*, 600(7889):547–552, Dec 2021.
- [77] Jue Wang, Sidney Lianza, David Juergens, Doug Tischer, Joseph L. Watson, Karla M. Castro, Robert Ragotte, Amijai Saragovi, Lukas F. Milles, Minkyung Baek, Ivan Anishchenko, Wei Yang, Derrick R. Hicks, Marc Expòsit, Thomas Schlichthaerle, Jung-Ho Chun, Justas Dauparas, Nathaniel Bennett, Basile I. M. Wicky, Andrew Muenks, Frank DiMaio, Bruno Correia, Sergey Ovchinnikov, and David Baker. Scaffolding protein functional sites using deep learning. *Science*, 377(6604):387–394, 2022.
- [78] BIM Wicky, LF Milles, A Courbet, RJ Ragotte, J Dauparas, E Kinfu, S Tipps, RD Kibler, M Baek, F DiMaio, et al. Hallucinating symmetric protein assemblies. *Science*, 378(6615):56–61, 2022.
- [79] Namrata Anand, Raphael Eguchi, Irimpan I. Mathews, Carla P. Perez, Alexander Derry, Russ B. Altman, and Po-Ssu Huang. Protein sequence design with a learned potential. *Nature Communications*, 13(1):746, Feb 2022.

- [80] J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker. Robust deep learning based protein sequence design using proteinmpnn. *bioRxiv*, 2022.
- [81] Sunghwan Choi, Yeonjoon Kim, Jin Woo Kim, Zeehyo Kim, and Woo Youn Kim. Feasibility of activation energy prediction of gas-phase reactions by machine learning. *Chemistry–A European Journal*, 24(47):12354–12358, 2018.
- [82] Pascal Friederich, Gabriel dos Passos Gomes, Riccardo De Bin, Alán Aspuru-Guzik, and David Balcells. Machine learning dihydrogen activation in the chemical space surrounding vaska’s complex. *Chemical science*, 11(18):4584–4601, 2020.
- [83] Colin A Grambow, Lagnajit Pattanaik, and William H Green. Deep learning of activation energies. *The journal of physical chemistry letters*, 11(8):2992–2997, 2020.
- [84] Xin Li, Shuo-Qing Zhang, Li-Cheng Xu, and Xin Hong. Predicting regioselectivity in radical c- h functionalization of heterocycles through machine learning. *Angewandte Chemie International Edition*, 59(32):13253–13259, 2020.
- [85] Ignacio Migliaro and Thomas R Cundari. Density functional study of methane activation by frustrated lewis pairs with group 13 trihalides and group 15 pentahalides and a machine learning analysis of their barrier heights. *Journal of Chemical Information and Modeling*, 60(10):4958–4966, 2020.
- [86] Koichiro Mikami. Interactive-quantum-chemical-descriptors enabling accurate prediction of an activation energy through machine learning. *Polymer*, 203:122738, 2020.
- [87] Ferruccio Palazzesi, Markus R Hermann, Marc A Grundl, Alexander Pautsch, Daniel Seeliger, Christofer S Tautermann, and Alexander Weber. Bireactive: a machine-learning model to estimate covalent warhead reactivity. *Journal of Chemical Information and Modeling*, 60(6):2915–2923, 2020.
- [88] Kjell Jorner, Tore Brinck, Per-Ola Norrby, and David Buttar. Machine learning meets mechanistic modelling for accurate prediction of experimental activation energies. *Chemical Science*, 12(3):1163–1175, 2021.
- [89] Elliot H. E. Farrar and Matthew N. Grayson. Machine learning and semi-empirical calculations: a synergistic approach to rapid, accurate, and mechanism-based reaction barrier prediction. *Chem. Sci.*, 13:7594–7603, 2022.
- [90] Ryo Nagai, Ryosuke Akashi, and Osamu Sugino. Completing density functional theory by machine learning hidden messages from molecules. *npj Computational Materials*, 6(1):1–8, 2020.
- [91] Roman Zubatyuk, Justin S Smith, Jerzy Leszczynski, and Olexandr Isayev. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Science advances*, 5(8):eaav6490, 2019.

- [92] Andrzej M Zuranski, Jesus I Martinez Alvarado, Benjamin J Shields, and Abigail G Doyle. Predicting reaction yields via supervised learning. *Accounts of chemical research*, 54(8):1856–1865, 2021.
- [93] Benjamin J Shields, Jason Stevens, Jun Li, Marvin Parasram, Farhan Damani, Jesus I Martinez Alvarado, Jacob M Janey, Ryan P Adams, and Abigail G Doyle. Bayesian reaction optimization as a tool for chemical synthesis. *Nature*, 590(7844):89–96, 2021.
- [94] Lin Jiang, Eric A Althoff, Fernando R Clemente, Lindsey Doyle, Daniela Rothlisberger, Alexandre Zanghellini, Jasmine L Gallaher, Jamie L Betker, Fujie Tanaka, Carlos F Barbas III, et al. De novo computational design of retro-aldol enzymes. *science*, 319(5868):1387–1391, 2008.
- [95] Daniela Röthlisberger, Olga Khersonsky, Andrew M Wollacott, Lin Jiang, Jason DeChancie, Jamie Betker, Jasmine L Gallaher, Eric A Althoff, Alexandre Zanghellini, Orly Dym, et al. Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192):190–195, 2008.
- [96] Hsien-Wei Yeh, Omran Karmach, Ao Ji, David Carter, Manuela M Martins-Green, and Huiwang Ai. Red-shifted luciferase–luciferin pairs for enhanced bioluminescence imaging. *Nature methods*, 14(10):971–974, 2017.

Chapter 2

Enantioselective Diarylcarbene Insertion into Si–H Bonds Induced by Electronic Properties of the Carbenes

2.1 Author List and Affiliation

Liang-Liang Yang¹, Declan Evans², Bin Xu¹, Wen-Tao Li¹, Mao-Lin Li¹, Shou-Fei Zhu^{1,*}, K. N. Houk^{2,*}, and Qi-Lin Zhou^{1,*}

¹State Key Laboratory and Institute of Elemento-Organic Chemistry, College of Chemistry, Nankai University, Tianjin 300071, China

²Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095-1569, United States

*Corresponding author

This work is a collaboration between the groups of Qi-Lin Zhou (Nankai) and K. N. Houk (UCLA), published in the Journal of the American Chemical Society at *J. Am. Chem. Soc.* **2020**, 142 (28), 12394-12399. Experiments were led by Liang-Liang Yang under the guidance of Qi-Lin Zhou. DFT Calculations were performed by Declan Evans under the guidance of K. N. Houk.

2.2 Abstract

Catalytic enantioselection usually depends on differences in steric interactions between prochiral substrates and a chiral catalyst. We have discovered a carbene Si–H insertion in which the enantioselectivity depends primarily on the electronic characteristics of the carbene substrate, and the log(*er*) values are linearly related to Hammett parameters. A new class of chiral tetrakisphosphate dirhodium catalysts was developed; it shows excellent activity and enantioselectivity for the insertion of diarylcarbenes into the Si–H bond of silanes. Computational and mechanistic studies show how the electronic differences between the two aryls of the carbene lead to differences in energies of the diastereomeric transition states. This study provides a new strategy for asymmetric catalysis exploiting the electronic properties of the substrates.

2.3 Introduction

Enantiomers of chiral molecules often exhibit biological activities distinct from one another because of the homochirality of biological molecules (e.g., l-amino acids and d-saccharides). As a result, industrial compounds such as pharmaceuticals, pesticides, flavors, and fragrances often must be a single enantiomer [1]. In the past several decades, asymmetric catalysis has become a reliable method for the synthesis of enantiomerically enriched chiral compounds [2]. Most current methods of asymmetric catalysis rely on the spatial interactions between the catalyst and substrate for enantiocontrol. Many sterically crowded chiral catalysts have been developed that utilize this principle to achieve high enantioselectivity [3, 4, 5]. For unsaturated substrates, the chiral catalyst is often able to achieve enantiocontrol by discriminating between the two prochiral faces of the substrate [6]. It is difficult to identify prochiral faces when the reactive center is attached to two sterically similar substituents (Figure 2.1A). For this reason, only limited successes have been achieved in the catalytic enantioselective reactions, such as hydrogenations of diaryl [7] or dialkyl ketones [8, 9] or of diaryl or dialkyl ethylenes [10, 11, 12] and cycloadditions of diaryl ethylenes [13].

Many active intermediates of organic reactions, such as carbocations, carbon radicals, carbenes, and conjugated carbanions, have planar structures. Asymmetric reactions that proceed via these active intermediates often rely on the steric differences between the substituents connected to the prochiral center [14, 15, 16, 17]. Catalytic enantiocontrol by discriminating substituent electronics is very rare [18, 19, 20, 21]. Recently, Fürstner and co-workers [22] studied the structures of rhodium-diphenylcarbenes by X-ray single-crystal diffraction: the electron-rich phenyl ring (4-Me₂NC₆H₄) adopts a coplanar orientation with the carbene plane ($\theta = 0.9^\circ$) to maximize the overlap of the π cloud of the phenyl ring with the (empty) carbene p orbital. The electron-deficient phenyl ring (4-CF₃C₆H₄) lies orthogonal to the carbene plane ($\theta = -94.7^\circ$) to stabilize the lone pair of the carbene that also donates to the Rh d_{z^2} orbital. In this way, the electronic properties of the substituents determine the degree of conjugation between the aryl rings and the p or

lone-pair orbitals of the carbene center. On the basis of this property, we speculated that enantioselective transformation could be achieved by a chiral catalyst that distinguishes the conformations of a prochiral carbene intermediate. This strategy provides a new method for enantiocontrol for chiral transformations of substrates, which have sterically similar substituents. Although stereoselectivity has been achieved in carbene reactions catalyzed by Cu, Fe, and Rh catalysts [23, 24], the carbenes generally have one ester group and one aryl or alkenyl group, or in the work by Shaw, very different substitution patterns on the two aryl groups of diaryl carbenes that cause large steric differences [25]. We have now found that dirhodium catalysts modified with chiral spiro phosphate ligands (Figure 2.1B, $\text{Rh}_2(\text{SPA})_4$) can differentiate the conformations of diarylcarbenes to achieve highly enantioselective transformations even though the substituents are in positions where they have no different direct interactions with the catalyst.

Transition-metal-catalyzed carbene insertion into Si–H bonds is a powerful method for the synthesis of optically active silanes. However, high enantioselectivity has been achieved only with the use of carbenes with markedly different substituents (e.g., one substituent is an alkyl, alkenyl, or aryl group and the other substituent is an electron-withdrawing group) [26, 27, 28, 29, 30]. Enantioselective Si–H bond insertion of diarylcarbenes remains undeveloped. The challenge in the enantioselective Si–H bond insertion of diarylcarbene stems from the difficulty of precisely distinguishing the Re and Si faces of the carbene. We sought to overcome this limitation using a chiral catalyst capable of distinguishing the electronic-induced conformations of the prochiral carbene intermediate (Figure 4.1B).

2.4 Results and Discussion

The insertion of 4-nitrophenyl phenyl diazomethane (**D1**) into dimethylphenylsilane (**S1**) was performed using several chiral dirhodium catalysts commonly used in asymmetric carbene transformations [31, 32, 33]. The best performing catalyst for this reaction, $\text{Rh}_2(\text{S-PTTL})_4$, gave high yield but only modest enantioselectivity, prompting further investigation. A new type of dirhodium

catalyst that contains spiro phosphate ligands C1–C5 was developed [34]. Of the spiro phosphate dirhodium catalysts, C2 afforded both good yield and the highest enantioselectivity (Figure 2.2). Even 0.01 mol % catalyst is sufficient for excellent results.

A variety of diphenyl diazomethylenes **D1–D12** bearing electronically different para substituents were evaluated in the reaction with silane **S1** catalyzed by C2 (Table 2.1). The substrates with a strong electron-withdrawing group or a strong electron-donating group at the para position of one phenyl ring exhibited high enantioselectivity (entries **1–4** and **12**), whereas the substrates having a moderate or weak electronic effect exhibited lower enantioselectivity (entries **5–11**). Moreover, the substrates with electron-withdrawing groups (**D3**, R = CF₃) or electron-donating groups (**D11**, R = OMe) afforded Si–H bond insertion products with opposite absolute configurations (entry **3** and entry **11**). When the Hammett substituent constant (σ_p) [35] differences of the para substituents of two phenyl rings are over 0.5, the e.e. values of the corresponding Si–H bond insertion products are over 90%. Moreover, a plot of log(er) values against Hammett's σ_p values is shown in Figure ?? (slope 2.86, R² = 0.96). These results clearly indicate that the enantioselectivity of this reaction is directly related to the electronic properties of carbene intermediates.

Next, various diphenyl diazomethanes bearing different para substituents were studied (Table 2.2). These results show that the enantioselectivity of the reaction is directly correlated with the differences in substrate electronics. Again, when the difference of the Hammett substituent constant of the two para substituents is equal or greater than 0.5, the e.e. of the corresponding Si–H bond insertion product is equal or greater than 90% (entries **1–6**, **9–11**, and **14–16**). Moreover, sterically similar substituents (e.g., *p*-NO₂ vs *p*-NMe₂, entry **1**; *p*-CF₃ vs *p*-CH₃, entry **10**; *p*-OCF₃ vs *p*-OCH₃, entry **14**) were precisely differentiated, indicating that the observed enantioselectivity is not a result of steric effects.

The chiral spiro phosphate dirhodium catalyst **C2** is also efficient for the enantioselective Si–H bond insertion of other diazo compounds containing two different aryl groups (Figure 2.3). In every case, the electronic property of the substituents on the aryl rings significantly affected the enantioselectivity of the reaction. For instance, the naphthyl phenyl diazomethane afforded the Si–H bond

insertion product **P31** with 53% e.e.; however, introducing a *p*-NO₂ at the phenyl ring dramatically increased the e.e. value of product **P32** to 96%. Similarly, 4-nitrophenyl 2'-methoxyphenyl diazomethane provided much higher enantioselectivity (**P34**, >99% e.e.) than 2-methoxyphenyl phenyl diazomethane (**P33**, 32% e.e.). Moreover, if one aryl ring of the substrates has a strong electron-withdrawing para substituent (e.g., NO₂ or CF₃) and the other aryl ring is a heteroaryl ring (**P36–P38**), excellent enantioselectivity can be obtained. Excellent enantioselectivity can also be achieved when the diaryl diazomethane substrates have an ortho substituent (**P39** and **P40**). Presumably because of the small radius of the fluorine atom, the 2-fluoro-substituted diazo compound afforded relatively lower e.e. (83% e.e., **P41**). However, the enantioselectivity can be increased by introducing an electron-donating group (4-OMe) at the other aryl ring of the substrate (91% e.e., **P42**). In addition to dimethylphenylsilane (**S1**), other silanes can also be used in the Si–H bond insertion reaction with diaryl diazomethane **D1**, affording Si–H bond insertion products with excellent enantioselectivity (**P43–P49**). It is worth mentioning that the alkynyl silane underwent Si–H bond insertion reaction, giving the corresponding product **P47** with high yield (93%) and excellent enantioselectivity (> 99% e.e.). The absolute configuration of (*S*)-**P47** was determined by single-crystal X-ray diffraction analysis.

2.5 Mechanistic Studies

A kinetic isotopic study was carried out; in a competition experiment using dimethylphenylsilane and deuterated dimethylphenylsilane, the kinetic isotope effect (KIE) was found to be 1.5. This result is similar to the values reported in the Si–H insertions of aryl diazoesters catalyzed by transition metals such as Ir [28], Rh [29], and Cu [36]; these reactions proceed through concerted three-center transition states. To gain insights about the origin of enantioselectivity from the catalyst and the electronic effects of substrates, we also conducted a computational investigation. **C1** was used as the model catalyst to avoid the conformational complexity from the relative rotation of the outermost phenyl of **C2**. X-ray structure analysis of the catalyst **C1** reveals that the presence

of four identical chiral ligands around the dirhodium core results in a rigid catalyst with higher symmetry (D_4 symmetry) than the ligands themselves (C_2 symmetry) [34]. This symmetry causes both sides of the dirhodium catalyst to be identical, limiting the number of possible conformations for the transition state of Si–H insertion.

Initially, a dirhodium-tetraformate catalyst ($\text{Rh}_2(\text{O}_2\text{CH})_4$) was computed to observe transition-state geometries in the absence of any steric effects caused by the ligands. Figure 2.4A shows that the Si–H insertion proceeds via a concerted three-center transition state. Calculated transition-state geometries for diphenylcarbene insertion into the Si–H bond of dimethylphenylsilane show that one aryl ring rotates to a near-coplanar orientation with the carbene empty p orbital, while the other phenyl ring rotates to a near orthogonal orientation. For substituted diphenylcarbenes, two transition-state geometries are possible, depending on whether the substituted aryl ring is conjugated (**TS-1**) or orthogonal (**TS-2**) to the empty p orbital. These two transition states are not equal in energy and the favored transition state always results when the electron-rich aryl ring is nearly coplanar with the carbene plane. Plotting the energy difference ($\Delta\Delta G^\ddagger$) vs the corresponding Hammett substituent constant σ_p shows a linear correlation (Figure 2.4A), indicating that this energy difference is greater when the electronic difference between aryl rings is more pronounced.

In the chiral environment of **C1**, the transition state for the Si–H bond approaching the asymmetrical diphenylcarbene ($\text{R}^1 = \text{NO}_2$, $\text{R}^2 = \text{OMe}$) from its Re face (**TS-3**) is the lowest in energy (Figure 2.4B,C). This transition state is favored by 5.6 kcal mol⁻¹ over **TS-5** because of the different steric effects of the two aryl substituents on carbene; this steric difference originates from the electronic effects of substituents that orient electron-rich and electron-deficient aryl rings differently. Moreover, **TS-3** is favored by 2.2 kcal mol⁻¹ over **TS-4** (which would lead to the opposite enantiomer) because of the steric effects that the chiral catalyst has on the transition-state conformations. In these structures, the electron-rich aryl ring ($\text{R}^2 = \text{OMe}$) is able to adopt a more coplanar orientation in **TS-3** ($\theta = 19.5^\circ$) than in **TS-4** ($\theta = 31.6^\circ$). The last possible transition state, **TS-6**, is 7.4 kcal mol⁻¹ higher in energy than **TS-3** because of both unfavorable electronics and steric clashes with the catalyst. Moreover, when two aryls of the carbene have similar electronic

properties, as in the case for $R^1 = \text{CN}$ and $R^2 = \text{CF}_3$, the energy difference between **TS'-3** and **TS'-5** is smaller ($2.4 \text{ kcal mol}^{-1}$), which is consistent with the low enantioselectivity observed in the experiment [37, 38, 39, 40, 41].

These calculations show that the chiral environment of the catalyst **C1** can fix the conformations of the transition state in the Si–H bond insertion reaction. On this basis, the energy difference of the transition states, which relates to the enantioselectivity of the reaction, is primarily due to the electronic difference between two aryls of carbene.

2.6 Conclusion

In summary, we have achieved a method for highly enantioselective carbene insertion into Si–H bonds that relies primarily on the electronic properties of the substrates. It represents the first highly enantioselective diarylcarbene insertion into the heteroatom–hydrogen bonds. A new class of D_4 symmetric dirhodium catalysts bearing chiral spiro phosphate ligands was developed, and computational studies demonstrate that the chiral environment of these catalysts can differentiate the various possible transition-state conformations. The chiral induction observed in this study not only enables the unprecedented enantioselective transformations of carbenes having spatially similar groups but also inspires the development of new strategies for chiral transformations of charged active intermediates, such as carbocations, carbanions, and carbon radicals.

2.7 Methods

2.7.1 General Information

NMR spectra were recorded with a Bruker AV 400 spectrometer at 400 MHz (^1H NMR), 101 MHz (^{13}C NMR), 79 MHz (^{29}Si NMR), 376 MHz (^{19}F NMR), 128 MHz (^{11}B NMR) and 162 MHz (^{31}P NMR). Chemical shifts (δ values) were reported in ppm down field from internal Me_4Si (^1H and ^{13}C NMR). High Resolution Mass Spectra (HRMS) were recorded on an IonSpec FT-ICR

mass spectrometer with Electron Spray Ionization (ESI) resource and Waters GCT Premier mass spectrometer. Melting points were measured on a RY-I apparatus and uncorrected. Enantioselectivities were recorded on Agilent HPLC, using a chiral stationary phase column (Daicel Co. CHIRALPAK, CHIRALCEL, Phenomenex). The chiral HPLC methods were calibrated with the corresponding racemic mixtures. All reactions and manipulations were performed using standard Schlenk techniques. All solvents were purified and dried using standard procedures. All chiral carboxylate dirhodium complexes were purchased from Strem and TCI.

2.7.2 Preparation of substrates

Preparation of diarylmethanones

(4-Nitrophenyl)(phenyl)methanone (**K1**), 4-benzoylbenzotrile (**K2**), phenyl(4-(trifluoromethyl)phenyl)methanone (**K3**), (4-chlorophenyl)(phenyl)methanone (**K6**), (4-bromophenyl)(phenyl)methanone (**K7**), (4-iodophenyl)(phenyl)methanone (**K8**), (4-fluorophenyl)(phenyl)methanone (**K9**), phenyl(p-tolyl)methanone (**K10**), (4-methoxyphenyl)(phenyl)methanone (**K11**), (4-(dimethylamino)phenyl)(phenyl)methanone (**K12**), (4-methoxyphenyl)(4-nitrophenyl)methanone (**K14**), (4-methoxyphenyl)(4-(trifluoromethyl)phenyl)methanone (**K26**), (4-bromophenyl)(4-methoxyphenyl)methanone (**K27**), (4-chlorophenyl)(4-methoxyphenyl)methanone (**K28**), (4-fluorophenyl)(4-methoxyphenyl)methanone (**K29**), (4-methoxyphenyl)(p-tolyl)methanone (**K30**), naphthalen-2-yl(phenyl)methanone (**K31**), (2-methoxyphenyl)(phenyl)methanone (**K33**), (2-chlorophenyl)(phenyl)methanone (**K39**), (2-bromophenyl)(phenyl)methanone (**K40**), (2-fluorophenyl)(phenyl)methanone (**K41**) and (2-fluorophenyl)(4-methoxyphenyl)methanone (**K42**) were purchased from Aldrich, Alfa, or TCI chemical company.

The diarylmethanones **K41**, **K52**, **K13** [42], **K15-K18** [43], **K19** [44], **K20** [45], **K21-K25** [44, 45], **K32** [46], **K34-K35** [46], **K36** [47] were prepared according to literature methods. The new diarylmethanones were prepared by using the following procedure (taking **K37** for example).

Step 1: A dry 250 mL Schlenk-tube equipped with a magnetic stir bar, was charged with benzofuran-

5-carbaldehyde (1.5 g, 10.0 mmol) and THF (25 mL) under argon. Under vigorous stirring, a solution of bromo[4-(trifluoromethyl)phenyl]magnesium (15.0 mmol) in THF (40 mL) was slowly added under 0 °C. The reaction mixture was stirred under argon for 1.0 h and quenched with saturated aqueous NH₄Cl (20 mL). The aqueous layer was extracted with diethyl ether (3 x 20 mL). The combined organic extract was washed with brine, then dried over MgSO₄, filtered, and concentrated to give crude product. The crude product was purified by silica gel column chromatography (eluting with petroleum ether/EtOAc, 8:1, v/v) to afford benzofuran-5-yl(4-(trifluoromethyl)phenyl)methanol (2.5 g, 85% yield) as an oil.

Step 2: A 100 mL Schlenk-tube equipped with a magnetic stir bar, was charged with benzofuran-5-yl(4-(trifluoromethyl)phenyl)methanol (2.3 g, 7.8 mmol) and THF (30 mL). Under vigorous stirring, 2-Iodoxybenzoic acid (3.3 g, 11.7 mmol) was added at one portion at ambient temperature. The reaction mixture was stirred for 2 h and diluted by water. The mixture was filtered and washed with EtOAc. The aqueous layer was extracted with EtOAc (3 x 40 mL). The combined organic extract was washed with brine, then dried over MgSO₄, filtered, and concentrated to give a solid. The crude product was purified by silica gel column chromatography (eluting with petroleum ether/EtOAc, 15:1, v/v) to afford benzofuran-5-yl(4-(trifluoromethyl)phenyl)methanone (**K37**) (2.0 g, 86% yield) as a solid.

Preparation of diaryl hydrazones and diazomethanes

Diaryl hydrazones and diazomethanes were prepared according to the literature methods [48]. The following are the typical procedure for the preparation of diaryl hydrazones and diazomethanes.

Step 1: Hydrazine monohydrate (80% purity, 18 mL, 300.0 mmol) was added to (4-nitrophenyl)(phenyl)methanone (6.8 g, 30.0 mmol) in ethanol (100 mL). HOAc (0.5 mL) was added and the mixture was stirred at reflux for 12 h. After cooling to room temperature, (4-nitrophenyl)(phenyl)methanone hydrazone precipitated as yellow crystal. Filtration of the crude mixture gave pure (4-nitrophenyl)(phenyl)methanone

hydrazone as a yellow solid in 70% yield (5.0 g, 21.0 mmol). When the hydrazone could not precipitate. The EtOH was removed in vacuo, the residue was taken up in Et₂O and water. The layers were separated and the aqueous layer was extracted with Et₂O (3 x 20 mL). The combined organic layer was dried with MgSO₄, filtered, then concentrated in vacuo. Column chromatography using 15–25% EtOAc/hexanes as the eluent afforded the desired product. (Note: An inseparable mixture of two isomers of ketone hydrazone was obtained, which was used in the next step without further purification.)

Step 2: (4-Nitrophenyl)(phenyl)methanone hydrazone (2.9 g, 12.0 mmol) and anhydrous MgSO₄ (3.6 g) were placed into an oven-dried Schlenk tube under argon atmosphere and 50 mL DCM was introduced. After cooling to 0 °C, to this rapidly stirring mixture was added activated MnO₂ (3.7 g, 42.0 mmol) in one portion. The reaction mixture was warmed to room temperature and stirred for 5-6 h, and filtered off the solid. The solvent was removed under reduced pressure and the crude product was used for next reaction. (Note: The purity of compounds were determined by ¹H NMR, which was generally 90% 100%. The impurity was identified as (Ar¹Ar²)C=N-N=C(Ar¹Ar²).)

Preparation of silanes

Dimethyl(phenyl)silane (**S1**), ethoxydimethylsilane (**S2**), chlorodimethylsilane (**S3**), triethylsilane (**S4**), methyldiphenylsilane (**S5**) were purchased from Aldrich, Alfa, or TCI chemical company. Dimethyl(phenylethynyl)silane (**S7**) [49], (3-chloropropyl)dimethylsilane (**S8**) [50], 5-(dimethylsilyl)-1-methyl-1H-indole (**S9**) [49] were prepared according to the methodology described by the literatures.

2.7.3 Preparation of catalysts

Procedure for synthesis of C1[51]

According to our previously reported procedure, **SI-1** (3.4 g, 10.8 mmol), Rh₂(OAc)₄ (0.6 g, 1.4

mmol) and 130 mL dry toluene were introduced into a 250 mL two-necked flask. The mixture was refluxed for 36 h at reflux using a Soxhlet apparatus filled with sand and Na₂CO₃ for the removal of acetic acid. The mixture was then concentrated under vacuum and purified by column chromatography on silica gel using DCM as mobile phase to give the crude product. The crude product was recrystallized in a mixed solvent of DCM and MeOH to give **C1** as a green needle crystal (1.5 g, 93% yield).

Procedure for synthesis of C2 and C3

(S)-7,7'-dimethoxy-2,2',3,3'-tetrahydro-1,1'-spirobi[indene] (**SI-2**) was prepared from the optically pure (S)-SPINOL according to our previously reported procedures [52].

Step 1. Synthesis of SI-3 [53].

According to the reported procedure, in an oven-dried Schlenk tube (120 mL in volume) equipped with a stir bar were placed with **SI-2**, B2pin2 (4.5 g, 17.6 mmol), [Ir(OMe)(cod)]₂ (212 mg, 0.32 mmol) and dipyrindinyl tetraaminodiborane (248 mg, 0.64 mmol). After evacuation and refill with dry nitrogen for three times, dry methoxycyclopentane (30 mL) was added with syringe. The resulting mixture was stirred at 110 °C for 36 h. After cooling to room temperature, the reaction mixture was concentrated and the residue was purified by column chromatography on silica gel (eluting with petroleum ether/DCM = 1:1, v/v) to give **SI-3** as a white solid (3.8 g, 89% yield).

Step 2. Synthesis of SI-4

To a solution of SI-3 (1.3 g, 2.5 mmol), phenylboronic acid (2.4 g, 15 mmol) and Pd(PPh₃)₄ (433 mg, 0.375 mmol) in toluene (30 mL) and EtOH (14 mL), aqueous K₂CO₃ (1 M, 20 mL) was added. The mixture was degassed by freezing and then heated at reflux for 24 h. The reaction mixture was cooled to ambient temperature and diluted by ethyl acetate. The aqueous layer was extracted with EtOAc (3 x 20 mL). The combined organic extract was washed with brine, then dried over MgSO₄, filtered, concentrated. The residue was purified by column chromatography (eluting with petroleum ether/DCM = 4:1, v/v) to give **SI-4** as a white solid (0.8 g, 78% yield).

Step 3. Synthesis of SI-5

To a dried Schlenk tube equipped with septum and stir bar, **SI-4** (850 mg, 2.0 mmol) was added. After three vacuum nitrogen cycles, dry DCM (20 mL) was added by syringe. The solution was cooled to -78°C , treated with BBr_3 (1 M, 15 mmol) in DCM and allowed to warm to room temperature. After stirring overnight, the reaction mixture was diluted with DCM and washed sequentially with saturated NaHCO_3 and brine. The organic layer was dried over anhydrous MgSO_4 and then concentrated. The residue was purified by column chromatography (eluting with petroleum ether/DCM = 1:1, v/v) to give **SI-5** as a white solid (460 mg, 57% yield).

Step 4. Synthesis of **SI-6**

To a 50 mL oven-dried Schlenk flask containing **SI-5** (460 mg, 1.2 mmol) was added 4 mL anhydrous pyridine and freshly distilled POCl_3 (370 mg, 2.4 mmol) under a nitrogen atmosphere. The mixture was stirred under 90°C for 12 h. After cooling to room temperature, 4 mL of H_2O was added. The mixture was stirred under 90°C for another 12 h, then cooled by an ice-bath, followed by a slow addition of HCl (aq. 3 N, 35 mL). After stirring for 1 h, the mixture was extracted by DCM (3 x 20 mL). The combined organic layer was concentrated and the residue was purified by column chromatography (eluting first with petroleum ether/ EtOAc = 1:1, v/v, then MeOH/DCM = 1:5, v/v) to give 440 mg white solid. The white solid was dissolved with 100 mL DCM, washed with concentrated HCl (10 mL). The organic layer was dried by anhydrous MgSO_4 and then concentrated. The residue was dried under vacuum to give **SI-6** as a white solid (430 mg, 79% yield).

Step 5. Synthesis of **C2**

By using the same procedure as for **C1**. The catalyst **C2** was prepared in 65% yield. Green sheet crystal. m.p.: decomposed over 320°C . $[\alpha]_{\text{D}}^{25} +289$ (c 1.0, CHCl_3). ^1H NMR (400 MHz, CDCl_3) δ 7.64 (s, 8H, ArH), 7.42 – 7.37 (m, 16H, ArH), 7.33 (s, 8H, ArH), 7.22 – 7.13 (m, 24H, ArH), 3.17 – 3.06 (m, 8H, 4CH₂), 2.94 (s, 3H, H_2O), 2.92 – 2.80 (m, 8H, 4CH₂), 2.33 – 2.24 (m, 8H, 4CH₂), 2.07 – 1.95 (m, 13H, 4CH₂ + $(\text{CH}_3)_2(\text{CO})$). ^{13}C NMR (101 MHz, CDCl_3) δ 146.6 (p, $J = 3.3$ Hz), 146.3, 141.7, 140.2, 138.4, 128.5, 127.1, 127.0, 120.8, 120.8, 58.6, 38.6, 30.9 ($(\text{CH}_3)_2(\text{CO})$), 30.7. ^{31}P NMR (162 MHz, CDCl_3) δ 6.6 (s). HRMS (MALDI) calcd for

[C116H88O16P₄Rh₂, M + Na]⁺: 2090.3064, Found: 2090.3034.

Procedure for synthesis of C4 and C5

(S)-4,4'-dibromo-2,2',3,3'-tetrahydro-1,1'-spirobi[indene]-7,7'-diol (**SI-10**) and (S)-4,4'-diphenyl-2,2',3,3'-tetrahydro-1,1'-spirobi[indene]-7,7'-diol (**SI-11**) were prepared from the optically pure (S)-SPINOL according to our previously reported procedures [52].

2.7.4 Typical procedure for Si-H bond insertion reaction

General procedure A

Dimethyl(phenyl)silane **S1** (32.7 mg, 0.24 mmol) and **C2** (0.42 mg, 0.0002 mmol, in 2 mL DCM) was injected into an oven-dried Schlenk tube under argon atmosphere. The mixture was cooled to 0 °C under ice-bath, a solution of **D1** (47.9 mg, 0.2 mmol, in 1 mL DCM) was introduced by syringe in 3 min. The color of diazo compound immediately disappeared after the addition. The TLC showed that the reaction completed as soon as the addition finished. Then the reaction mixture was concentrated and purified by a flash chromatography on silica gel (eluting with petroleum ether/EtOAc = 50:1, v/v) to give **P1** as a colorless oil (64.0 mg, 92% yield).

Taking P3 for example:

The **P3** (68 mg, 0.18 mmol) was added into a 25 mL oven-dried Schlenk tube under argon atmosphere, and 2 mL of DCM was injected. The mixture was cooled to 0 °C under ice-bath, a solution of BF₃/AcOH (28 μL, 143 μL/mmol) was introduced. The mixture was stirred at reflux for 4 h, and cooled to room temperature and quenched with saturated aqueous NaHCO₃ (2 mL). The mixture was extracted with Et₂O for three times and the combined organic extract was washed with brine, dried over MgSO₄, filtered, and concentrated to give a solid. The solid was dissolved in Et₂O (2 mL) and was successively added 3-chloroperoxybenzoic acid (155 mg, 0.9 mmol) and triethylamine (20 mg, 0.20 mmol) under ice-bath. The mixture was stirred at room temperature for 20 h and then diluted by Et₂O (2 mL). The mixture was successively washed by saturated aque-

ous NaHSO₃ and saturated NaHCO₃. The water layer was extracted with Et₂O for three times. The combined extract was washed with brine, dried over MgSO₄, filtered, and concentrated to give a solid. The crude product was purified by silica gel column chromatography (petroleum ether/EtOAc, 10:1, v/v) to afford alcohol **3ol** (34 mg, 75% yield) as a pale solid.

General procedure B

Hydrazone (0.2 mmol, 1 equiv) and anhydrous MgSO₄ (70 mg, 350 mg/mmol) were placed into an oven-dried Schlenk tube under argon atmosphere and 2 mL DCM was introduced. After cooling to 0 °C, to this rapidly stirring mixture was added activated MnO₂ (0.7 mmol, 3.5 equiv) in one portion. The reaction mixture was warmed to room temperature and kept stirring for 5-6 h, TLC indicated that the hydrazone was consumed completely. Then the solid was filtered off and washed with anhydrous DCM. The solution was concentrated to 1 mL, and was reacted with silane as in the procedure A.

Taking P29 for example:

The **P29** (35 mg, 0.1 mmol) was fluorinated and then treated as in Method A. The **P29** was dissolved in THF/MeOH (2 mL, 1:1, v/v), and was successively added KHCO₃ (100 mg, 1.0 mmol, 10 equiv) and KF (58 mg, 1.0 mmol, 10 equiv) under ice-bath. After stirring for several minutes, the mixture was added aqueous H₂O₂ (30%, 10 equiv). The resulting mixture was stirred at room temperature for 24 h, and extracted with Et₂O for three times. The combined extract was washed with brine, dried over MgSO₄, filtered, and concentrated to give a solid. The solid was purified by silica gel column chromatography (eluting with petroleum ether/EtOAc, 10:1, v/v) to afford alcohol **29ol** (10 mg, 45% yield).

Notes: For the Si-H insertion products **P3-P5**, **P9**, **P22-P25** and **P29**, the enantiomeric excesses (ee) cannot be directly measured, but determined after converting to the corresponding alcohols [54].

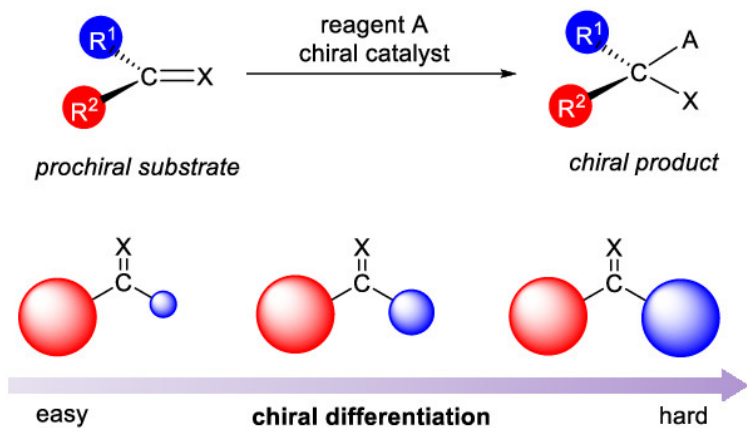
2.7.5 DFT Calculations

Conformations of each species were generated using Maestro with the OPLS3 force field. Conformational searches were conducted by restraining the position of the rhodium and rhodium-

bound atoms. Geometry optimizations and frequency calculations were performed with Gaussian09 [55] at the B3LYP-D3/6-31G(d)-LANL2DZ(Rh,I) level of theory [56]. Multi-layered ONIOM calculations [57] were performed on **C1** due to the size of the catalyst. Carbon and hydrogen atoms of **C1** were modeled with the Universal Force Field, and all other atoms with B3LYP-D3/6-31G(d)-LANL2DZ(Rh,I). Single point energy corrections were performed at the M06/6-311++G(d,p)-SDD(Rh,I)/SMD(DCM) level of theory [58]. Free energies were corrected with Truhlar's rigid-rotor harmonic oscillator treatment [59, 60] through Goodvibes [61].

Figure 2.1: Chiral differentiations of prochiral faces and enantioselective diarylcarbene insertion into Si-H bonds

A) Chiral differentiation via spatial interactions of substrate and catalyst



B) Enantioselective Si-H bond insertion induced by electronic properties of the diarylcarbenes

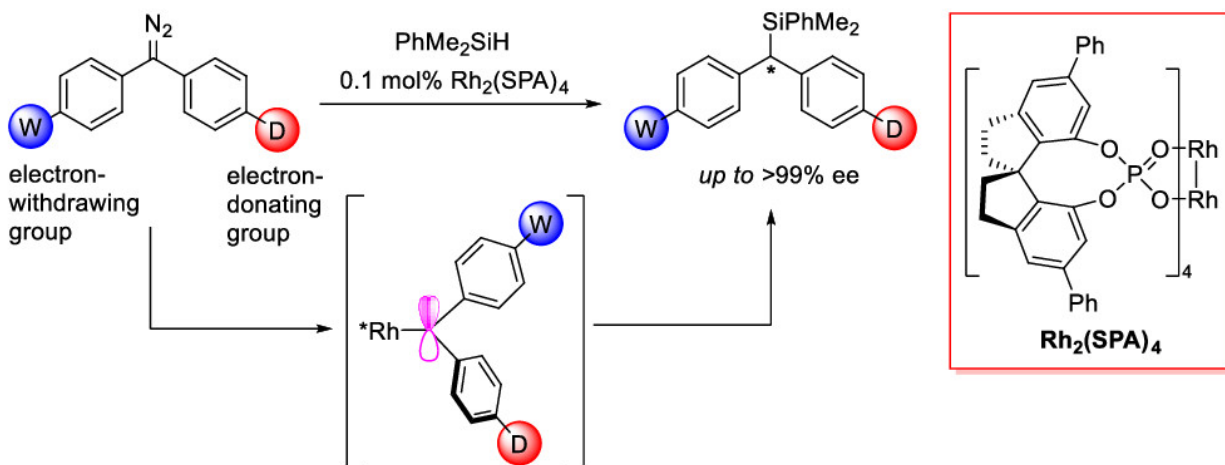
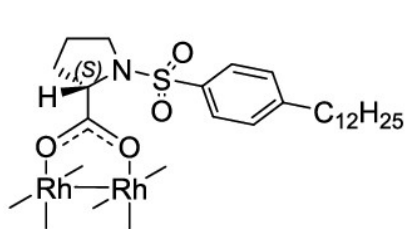
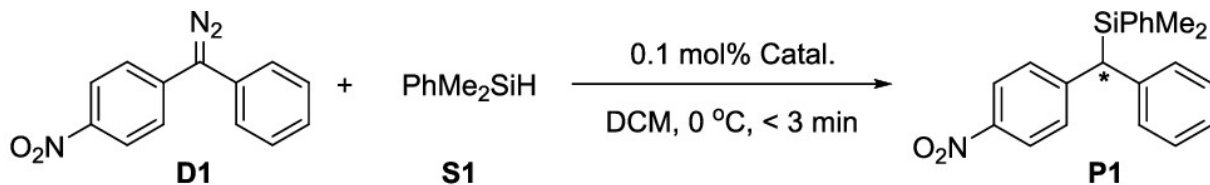
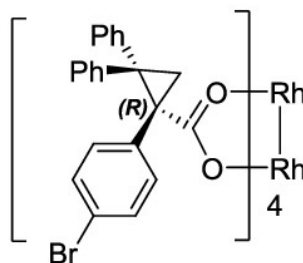


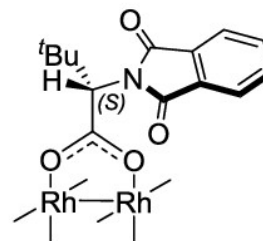
Figure 2.2: Enantioselective Si–H Bond Insertion of 4-Nitrophenylphenyl Diazomethane Catalyzed by Chiral Dirhodium Catalysts^a



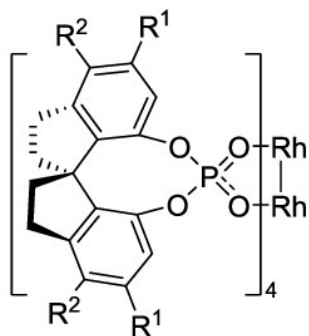
$\text{Rh}_2(\text{S-DOSP})_4$
85% yield, 17% ee



$\text{Rh}_2(\text{R-BTPCP})_4$
trace



$\text{Rh}_2(\text{S-PTTL})_4$
92% yield, 83% ee



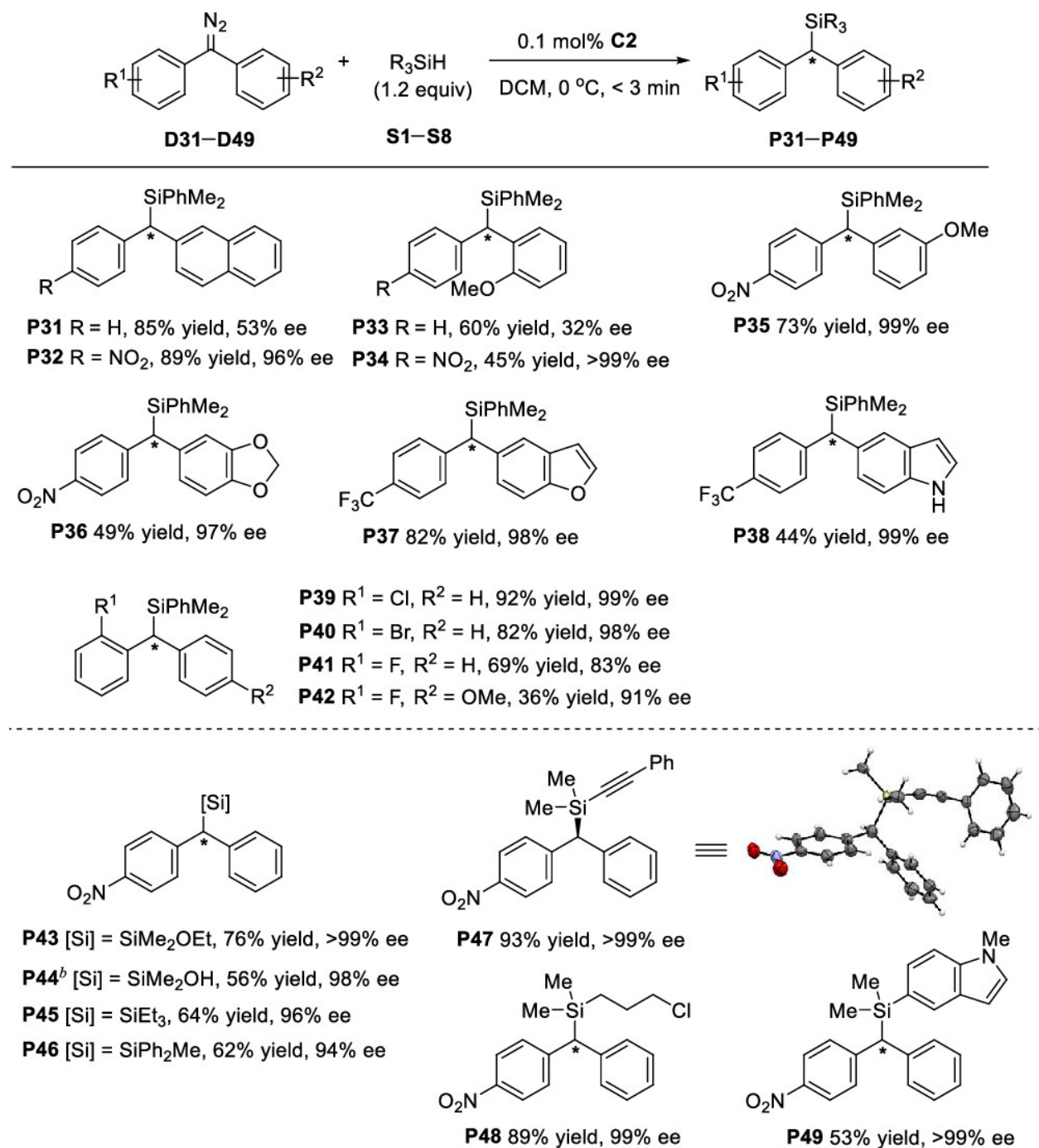
$\text{Rh}_2(\text{SPA})_4$

C1	$\text{R}^1 = \text{H}$	$\text{R}^2 = \text{H}$	56% yield 95% ee
C2	$\text{R}^1 = \text{C}_6\text{H}_5$	$\text{R}^2 = \text{H}$	92% yield >99% ee
C3	$\text{R}^1 = 3,5\text{-Me}_2\text{C}_6\text{H}_3$	$\text{R}^2 = \text{H}$	94% yield 97% ee
C4	$\text{R}^1 = \text{H}$	$\text{R}^2 = \text{Br}$	56% yield 66% ee
C5	$\text{R}^1 = \text{H}$	$\text{R}^2 = \text{C}_6\text{H}_5$	33% yield 82% ee

84% yield and >99% ee were obtained using 0.01 mol% **C2**!

^aReaction conditions: Catal/**D1**/**S1** = 0.0002:0.2:0.24 (mmol), 1 mL solution of **D1** was dropped into a 2 mL solution of **S1** and catalyst at 0 °C, 3 min, isolated yield.

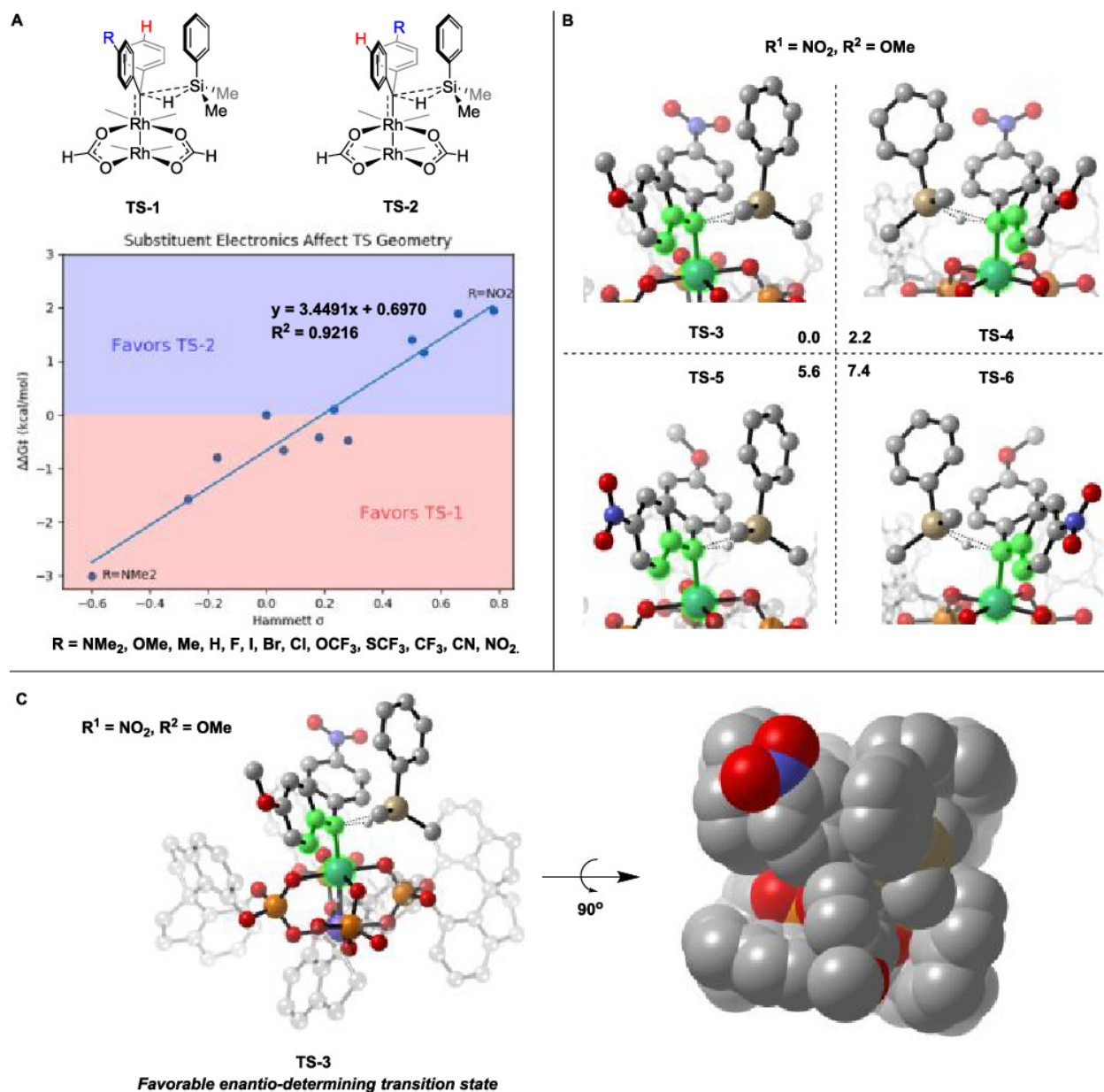
Figure 2.3: Substrate Scope of Diarylcarbene Insertion into Si–H Bonds^a



^aReaction conditions: hydrazone (0.2 mmol), MnO₂ (3.5 equiv), MgSO₄, DCM, rt, 5–6 h; then PhMe₂SiH (1.2 equiv), **C2** (0.1 mol %), DCM, 0 °C. Isolated yields were given. The e.e. values were determined by chiral HPLC analysis.

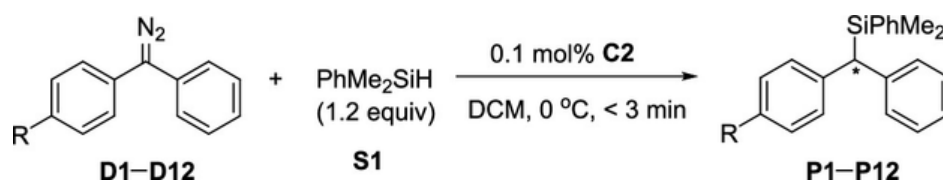
^bDimethylchlorosilane was used and the hydrolysis product was isolated after workup.

Figure 2.4: Computational studies



(A) Optimized transition-state structures for carbene insertion into the Si–H bond with a tetraformate dirhodium ($\text{Rh}_2(\text{O}_2\text{CH})_4$) as a model catalyst. (B) Structures of optimized transition states for the Si–H bond insertion of asymmetrically substituted diphenylcarbene formed from **C1**. ONIOM partitioning of the transition-state geometry, with the atoms shown opaque modeled with density functional theory (DFT), and the atoms shown transparent modeled with the universal force field (UFF). The relative Gibbs free energies with single-point corrections are given in kcal mol^{-1} . (C) Favorable enantio-determining transition-state structure for the Si–H bond insertion of asymmetrically substituted diphenylcarbene ($R^1 = \text{NO}_2, R^2 = \text{OMe}$) formed from **C1**.

Table 2.1: Rhodium-Catalyzed Enantioselective Si–H Bond Insertions of Diphenyl Diazomethylenes D1–D12



Entry ^a	R	$\Delta\sigma_p$ (R–H) ^b	Product	Yield (%)	e.e. (%)
1	NO ₂	0.78	P1	92	> 99
2	CN	0.66	P2	96	98
3 ^c	CF ₃	0.54	P3	91	95(<i>S</i>) ^d
4 ^c	SCF ₃	0.50	P4	68	97
5 ^c	OCF ₃	0.35	P5	66	76
6	Cl	0.28	P6	95	66
7	Br	0.23	P7	79	77
8	I	0.18	P8	62	82
9 ^c	F	0.06	P9	88	25
10	Me	-0.17	P10	51	18
11	OMe	-0.27	P11	47	64(<i>R</i>) ^d
12	NMe ₂	-0.60	P12	45	91

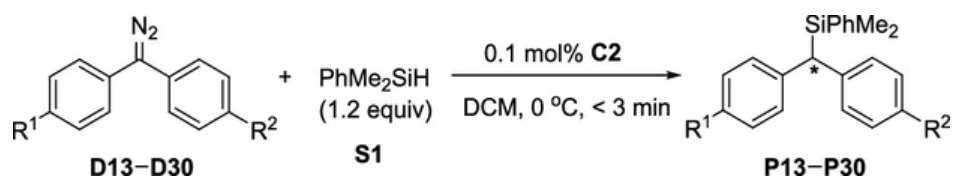
^aReaction conditions: Condition A (for entries **1–3**, **6**, **7**, **10**, and **11**): diazo compound (0.2 mmol), PhMe₂SiH (1.2 equiv), **C2** (0.1 mol %), DCM, 0 °C. Condition B (for entries **4**, **5**, **8**, **9**, and **12**): hydrazone (0.2 mmol), MnO₂ (3.5 equiv), MgSO₄, DCM, rt, 5–6 h; then PhMe₂SiH (1.2 equiv), **C2** (0.1 mol %), DCM, 0 °C. Isolated yields were given. The e.e. values were determined by chiral HPLC.

^bHammett substituent constant for para substituents.

^cThe e.e. value was determined by the corresponding alcohol obtained through the oxidation of the Si–H insertion product.

^dThe absolute configuration was assigned by analogy with the reported data.

Table 2.2: Rhodium-Catalyzed Enantioselective Si–H Bond Insertion of Diphenyl Diazomethanes D13–D30



Entry ^a	R ¹	R ²	$\Delta\sigma_p$ (R ¹ -R ²) ^b	Product	Yield (%)	e.e. (%)
1	NO ₂	NMe ₂	1.38	P13	66	96
2	NO ₂	OMe	1.05	P14	78	99
3	NO ₂	Me	0.95	P15	87	99
4	NO ₂	F	0.72	P16	86	98
5	NO ₂	Br	0.55	P17	80	95
6	NO ₂	Cl	0.5	P18	88	96
7	NO ₂	CF ₃	0.24	P19	80	86
8	CN	CF ₃	0.12	P20	42	36
9	CF ₃	OMe	0.81	P21	78	98
10 ^c	CF ₃	Me	0.71	P22	85	94(<i>S</i>)
11 ^c	CF ₃	F	0.48	P23	63	92
12 ^c	CF ₃	Br	0.31	P24	71	67
13 ^c	CF ₃	Cl	0.26	P25	73	76(<i>R</i>)
14	OCF ₃	OMe	0.62	P26	58	93
15	Br	OMe	0.50	P27	58	94
16	Cl	OMe	0.55	P28	59	93
17 ^c	F	OMe	0.33	P29	43	77(<i>R</i>)
18	Me	OMe	0.10	P30	40	56

^aReaction conditions: hydrazone (0.2 mmol), MnO₂ (3.5 equiv), MgSO₄, DCM, rt, 5–6 h; then PhMe₂SiH (1.2 equiv), **C2** (0.1 mol %), DCM, 0 °C. Isolated yields were given. The e.e. values were determined by chiral HPLC analysis.

^bThe Hammett substituent constant difference for para substituents.

^cThe e.e. value was determined by the corresponding alcohol obtained through the oxidation of Si–H insertion product.

2.8 Computational Models

Table S2.1: Computed Energies of Dirhodium-Tetraformate Transition States (Hartree)

Structure	Electronic Energy	Zero Point Energy	Enthalpy	Free Energy	qH Free Energy
TS-1 (R=Br)	-4653.6958	0.454312	-4653.2026	-4653.313	-4653.3048
TS-2 (R=Br)	-4653.6959	0.45421	-4653.2026	-4653.3141	-4653.3049
TS-1 (R=CF ₃)	-2417.3084	0.469088	-2416.7981	-2416.9143	-2416.9041
TS-2 (R=CF ₃)	-2417.3105	0.469279	-2416.8	-2416.9162	-2416.9059
TS-1 (R=Cl)	-2539.8707	0.454748	-2539.3773	-2539.4866	-2539.4786
TS-2 (R=Cl)	-2539.8701	0.454802	-2539.3767	-2539.4869	-2539.4778
TS-1 (R=CN)	-2172.4855	0.463051	-2171.9833	-2172.0937	-2172.0857
TS-2 (R=CN)	-2172.4885	0.462981	-2171.9862	-2172.0972	-2172.0887
TS-1 (R=F)	-2179.5172	0.456153	-2179.0229	-2179.1306	-2179.1232
TS-2 (R=F)	-2179.5162	0.456112	-2179.0218	-2179.1299	-2179.1221
TS-1 (R=I)	-2091.0628	0.453815	-2090.5699	-2090.681	-2090.6728
TS-2 (R=I)	-2091.0615	0.453516	-2090.5687	-2090.681	-2090.6721
TS-1 (R=Me)	-2119.5715	0.491766	-2119.0405	-2119.1513	-2119.1428
TS-2 (R=Me)	-2119.5704	0.491862	-2119.0393	-2119.1503	-2119.1416
TS-1 (R=NMe ₂)	-2214.1929	0.537649	-2213.6134	-2213.7297	-2213.7206
TS-2 (R=NMe ₂)	-2214.1878	0.537483	-2213.6084	-2213.7247	-2213.7158
TS-1 (R=NO ₂)	-2284.744	0.467082	-2284.2369	-2284.3494	-2284.3407
TS-2 (R=NO ₂)	-2284.7474	0.467329	-2284.2402	-2284.3529	-2284.3438
TS-1 (R=OCF ₃)	-2492.52	0.473789	-2492.0047	-2492.1196	-2492.1117
TS-2 (R=OCF ₃)	-2492.5183	0.473769	-2492.0027	-2492.1201	-2492.1099
TS-1 (R=OMe)	-2194.7734	0.496911	-2194.2364	-2194.3487	-2194.3403
TS-2 (R=OMe)	-2194.7707	0.496603	-2194.2341	-2194.3462	-2194.3378
TS-1 (R=SCF ₃)	-2815.4781	0.470362	-2814.9651	-2815.083	-2815.074
TS-2 (R=SCF ₃)	-2815.4801	0.469987	-2814.9673	-2815.0878	-2815.0762

Table S2.2: Computed Dnergies of (S)-C1 Transition States (Hartree)

Structure	Electronic Energy	Zero Point Energy	Enthalpy	Free Energy
TS-3	-6835.565466	-6833.9538	-6833.8602	-6834.0875
TS-4	-6835.565181	-6833.9536	-6833.8602	-6834.0839
TS-5	-6835.558894	-6833.9472	-6833.8537	-6834.0785
TS-6	-6835.556087	-6833.9446	-6833.8512	-6834.0757
TS-3'	-6945.8393	-6944.2597	-6944.1659	-6944.392
TS-4'	-6945.8371	-6944.2577	-6944.1639	-6944.3904
TS-5'	-6945.8351	-6944.2556	-6944.1617	-6944.3881
TS-6'	-6945.8368	-6944.2574	-6944.1638	-6944.3881

Cartesian Coordinates**TS-1 (R=Br)**

Rh -3.93328 -0.78749 -0.34488
 Rh -1.55776 -0.22373 0.18343
 O -4.44838 0.17503 1.42185
 O -2.27301 0.67891 1.90977
 O -3.65253 -2.55628 0.71237
 O -1.48797 -2.02974 1.21823
 O -1.02775 -1.15770 -1.60970
 O -3.19732 -1.69648 -2.08145
 O -1.86192 1.53234 -0.90381
 O -4.04080 1.02926 -1.35812
 C 0.41693 0.26004 0.65212
 C 0.54676 1.41763 1.55495
 C 0.07745 2.70006 1.20812
 C 1.02512 1.21544 2.86772
 C 0.12234 3.74535 2.12196
 H -0.32964 2.85806 0.21787
 C 1.03116 2.25857 3.79239
 H 1.36025 0.22823 3.16816
 C 0.59520 3.53077 3.42087
 H -0.22318 4.73187 1.82460
 H 1.38698 2.07533 4.80258
 H 0.61803 4.34780 4.13680
 C 1.53577 -0.70809 0.62724
 C 1.34567 -2.04363 0.21783
 C 2.85629 -0.28702 0.91083
 C 2.41622 -2.92780 0.12508

H 0.35246 -2.38793 -0.02643
 C 3.93606 -1.15610 0.80389
 H 3.04051 0.74499 1.18614
 C 3.70480 -2.47781 0.41577
 H 2.25561 -3.95527 -0.18171
 H 4.94419 -0.81499 1.01242
 C -3.52611 0.67019 2.12363
 H -3.83775 1.16851 3.05418
 C -3.01190 1.74799 -1.40800
 H -3.10217 2.69399 -1.96421
 C -2.53190 -2.75916 1.24550
 H -2.42396 -3.69466 1.81550
 C -1.96358 -1.66712 -2.31272
 H -1.62919 -2.14285 -3.24817
 Si 1.37909 1.31672 -2.08093
 H 0.62597 0.91701 -0.77247
 C 2.13922 -0.27906 -2.71288
 H 1.34103 -1.01555 -2.84667
 H 2.64085 -0.11937 -3.67525
 H 2.86485 -0.70004 -2.01057
 C 0.11694 2.04769 -3.26428
 H -0.34118 2.95376 -2.85738
 H 0.58655 2.29320 -4.22484
 H -0.68110 1.31939 -3.43522
 C 2.64251 2.56137 -1.48049
 C 2.26772 3.89272 -1.21543
 C 3.96903 2.18278 -1.19667

C 3.17674 4.80851 -0.68631
H 1.25141 4.22107 -1.42050
C 4.88302 3.09584 -0.66809
H 4.29312 1.16264 -1.38535
C 4.48674 4.41000 -0.40920
H 2.86417 5.83010 -0.48693
H 5.90285 2.78320 -0.45889
H 5.19588 5.12118 0.00609
Br 5.17718 -3.67961 0.26550

TS-2 (R=Br)

Rh -3.76317 1.00933 -0.51813
Rh -1.52215 0.15390 0.18734
O -3.05732 2.91377 -0.07470
O -1.00963 2.12876 0.56902
O -4.36650 0.77686 1.45791
O -2.30520 0.04269 2.11202
O -2.20971 -1.76487 -0.27700
O -4.26114 -0.98403 -0.90938
O -0.95243 0.36543 -1.81356
O -2.98683 1.18565 -2.44243
C 0.34986 -0.53978 0.78269
C 1.41980 0.46447 0.64988
C 1.77697 1.03083 -0.58934
C 2.02204 0.98718 1.81382
C 2.72511 2.04130 -0.67299
H 1.29569 0.67292 -1.48972
C 2.94514 2.02793 1.74354
H 1.74236 0.59405 2.78531
C 3.30362 2.53720 0.49739
H 3.00965 2.45017 -1.63624
H 3.38851 2.43053 2.64742
C 0.56913 -1.63888 1.74609
C -0.51117 -2.35388 2.30491
C 1.88224 -2.07157 2.05285
C -0.28549 -3.42881 3.16095
H -1.52120 -2.05811 2.06692
C 2.10300 -3.15731 2.89421
H 2.72818 -1.56926 1.59780
C 1.01774 -3.83585 3.45763
H -1.13125 -3.95886 3.58982
H 3.11949 -3.47735 3.10629
C -1.87836 3.03219 0.35347
H -1.54139 4.05639 0.57431
C -1.79913 0.83047 -2.64423

H -1.42876 0.92540 -3.67683
C -3.52109 0.37383 2.29651
H -3.86734 0.28895 3.33794
C -3.40370 -1.87993 -0.71270
H -3.71142 -2.91160 -0.94537
Si 0.91562 -2.55422 -1.49982
H 0.49851 -1.36837 -0.58248
C 0.31469 -4.09767 -0.61477
H -0.76322 -4.00803 -0.44943
H 0.50969 -4.99572 -1.21379
H 0.78980 -4.22205 0.36297
C 0.09079 -2.29159 -3.16714
H 0.43150 -1.36960 -3.64735
H 0.29665 -3.13349 -3.83945
H -0.99037 -2.21030 -3.02098
C 2.78282 -2.41161 -1.55190
C 3.41066 -1.45189 -2.36916
C 3.59770 -3.18977 -0.70681
C 4.79343 -1.27153 -2.34200
H 2.81306 -0.83190 -3.03347
C 4.98226 -3.01497 -0.67753
H 3.14700 -3.93624 -0.05835
C 5.58171 -2.05240 -1.49293
H 5.25612 -0.52163 -2.97801
H 5.59278 -3.62771 -0.01915
H 6.65893 -1.91099 -1.46754
H 1.18790 -4.68225 4.11783
Br 4.60232 3.93276 0.38584

TS-1 (R=CF3)

Rh -3.96278 -0.69943 -0.34293
Rh -1.57644 -0.17551 0.17754
O -4.47079 0.38240 1.35586
O -2.28742 0.86452 1.82831
O -3.74731 -2.40496 0.82581
O -1.57728 -1.90699 1.33554
O -1.05651 -1.25173 -1.53448
O -3.23892 -1.73246 -2.01158
O -1.81149 1.50239 -1.04152
O -4.00675 1.05324 -1.46937
C 0.40083 0.29085 0.64918
C 0.54881 1.49152 1.48969
C 0.05837 2.75047 1.08600
C 1.06451 1.36651 2.79781
C 0.11998 3.84537 1.93805

H -0.37756 2.84869 0.09960
C 1.08826 2.46013 3.66228
H 1.41624 0.40072 3.14478
C 0.63220 3.70654 3.23282
H -0.24510 4.81088 1.59790
H 1.47359 2.33632 4.67058
H 0.66807 4.56227 3.90144
C 1.50543 -0.69953 0.68276
C 1.28449 -2.05858 0.38609
C 2.83560 -0.27527 0.91218
C 2.34503 -2.95953 0.34351
H 0.28035 -2.40435 0.19478
C 3.89643 -1.16952 0.85063
H 3.03659 0.77142 1.10549
C 3.64976 -2.51653 0.56950
H 2.16192 -4.00658 0.12765
H 4.91208 -0.82558 1.01720
C -3.54189 0.90024 2.03174
H -3.84756 1.46596 2.92496
C -2.95141 1.72679 -1.56535
H -3.00316 2.63440 -2.18649
C -2.64203 -2.60408 1.39110
H -2.57038 -3.50306 2.02212
C -2.00094 -1.76730 -2.22075
H -1.67165 -2.31536 -3.11751
Si 1.47399 1.17028 -2.11499
H 0.66866 0.87305 -0.81158
C 2.03860 -0.50666 -2.74165
H 2.72171 -0.99376 -2.03877
H 1.16335 -1.15277 -2.85555
H 2.54516 -0.41718 -3.71053
C 0.31195 2.06487 -3.28977
H -0.04162 3.00621 -2.85786
H 0.81739 2.28503 -4.23813
H -0.56244 1.43877 -3.48851
C 2.89019 2.23790 -1.51362
C 2.65342 3.54444 -1.04402
C 4.20461 1.73836 -1.43567
C 3.68382 4.31841 -0.51058
H 1.65040 3.96135 -1.08241
C 5.24100 2.51185 -0.91029
H 4.42202 0.73069 -1.77995
C 4.98028 3.80196 -0.44217
H 3.47606 5.32072 -0.14572
H 6.24910 2.10814 -0.86311

H 5.78428 4.40318 -0.02591
C 4.81183 -3.46439 0.43758
F 4.44348 -4.75005 0.61615
F 5.78538 -3.18672 1.33327
F 5.37323 -3.37557 -0.79198

TS-2 (R=CF3)

Rh -3.74901 0.95642 -0.56219
Rh -1.50850 0.15496 0.19742
O -3.10439 2.87663 -0.09143
O -1.05461 2.13930 0.60040
O -4.39817 0.70025 1.39716
O -2.33722 0.01121 2.09962
O -2.13831 -1.77845 -0.28911
O -4.19051 -1.04506 -0.97304
O -0.89249 0.39082 -1.78599
O -2.93073 1.16206 -2.46331
C 0.35102 -0.48940 0.83708
C 1.41065 0.52401 0.69104
C 1.83811 1.01046 -0.55854
C 1.93088 1.12577 1.85719
C 2.78338 2.02311 -0.64117
H 1.41731 0.58987 -1.46079
C 2.84945 2.16708 1.77225
H 1.58981 0.78805 2.82995
C 3.29060 2.60738 0.52324
H 3.12149 2.37318 -1.61071
H 3.22830 2.63389 2.67548
C 0.58212 -1.57752 1.80031
C -0.48825 -2.33968 2.31785
C 1.90101 -1.95321 2.15788
C -0.24956 -3.40532 3.18089
H -1.49980 -2.08746 2.03976
C 2.13503 -3.02900 3.00757
H 2.74103 -1.41325 1.73608
C 1.05820 -3.75497 3.52757
H -1.08689 -3.97308 3.57650
H 3.15475 -3.30514 3.26139
H 1.23935 -4.59437 4.19371
C -1.93995 3.02251 0.36615
H -1.63349 4.05335 0.59957
C -1.72904 0.83839 -2.63590
H -1.33497 0.94855 -3.65801
C -3.56553 0.31168 2.25490
H -3.93582 0.21099 3.28651

C -3.31785 -1.92118 -0.75533
H -3.59472 -2.95935 -0.99709
Si 0.98162 -2.54047 -1.47326
H 0.53948 -1.37235 -0.57005
C 0.47208 -4.10097 -0.55875
H -0.61090 -4.08150 -0.40194
H 0.73087 -4.99820 -1.13409
H 0.94657 -4.17169 0.42499
C 0.11276 -2.33629 -3.12730
H 0.40609 -1.40680 -3.62412
H 0.33262 -3.17631 -3.79711
H -0.96721 -2.29264 -2.95663
C 2.84104 -2.32213 -1.57401
C 3.41801 -1.45194 -2.51880
C 3.69700 -2.93718 -0.63989
C 4.79017 -1.20043 -2.52891
H 2.78900 -0.95954 -3.25726
C 5.07047 -2.68800 -0.64451
H 3.28718 -3.61475 0.10443
C 5.61895 -1.81615 -1.58780
H 5.21315 -0.52366 -3.26663
H 5.71257 -3.17379 0.08584
H 6.68738 -1.61772 -1.59041
C 4.35722 3.66182 0.43234
F 4.29415 4.53433 1.46315
F 5.59760 3.11520 0.45749
F 4.26779 4.37393 -0.71351

TS-1 (R=Cl)

Rh -3.82494 0.45926 0.27539
Rh -1.39078 0.09345 -0.14102
O -4.19974 -0.92804 -1.22086
O -1.97374 -1.24010 -1.62715
O -3.73645 1.96808 -1.15253
O -1.50592 1.64024 -1.52479
O -1.03265 1.45156 1.40744
O -3.25905 1.82404 1.75066
O -1.48336 -1.41952 1.29430
O -3.71135 -1.09111 1.67008
C 0.64887 -0.23550 -0.54339
C 0.87894 -1.48003 -1.31062
C 0.55461 -2.74797 -0.78748
C 1.28568 -1.40491 -2.65813
C 0.68750 -3.89587 -1.55930
H 0.17322 -2.81538 0.22463

C 1.38742 -2.55644 -3.43929
H 1.49455 -0.43691 -3.10080
C 1.10761 -3.80786 -2.89151
H 0.44694 -4.86363 -1.12696
H 1.69452 -2.47053 -4.47812
H 1.20323 -4.70528 -3.49649
C 1.63091 0.86411 -0.71153
C 1.28385 2.20325 -0.43060
C 2.96983 0.60041 -1.08357
C 2.21731 3.22920 -0.53537
H 0.27491 2.43972 -0.13026
C 3.91320 1.61618 -1.18881
H 3.28690 -0.41897 -1.26511
C 3.52736 2.92755 -0.91297
H 1.93606 4.25533 -0.32449
H 4.93922 1.39054 -1.45620
C -3.21405 -1.43975 -1.81655
H -3.44881 -2.16304 -2.61236
C -2.60554 -1.65752 1.85156
H -2.58882 -2.47124 2.59396
C -2.63307 2.20254 -1.71033
H -2.62837 3.00509 -2.46367
C -2.03539 1.99647 1.97559
H -1.78073 2.71330 2.77224
Si 1.75952 -0.94180 2.18870
H 0.93620 -0.68447 0.85778
C 1.54902 0.64873 3.15654
H 0.49799 0.77822 3.42682
H 2.15868 0.64447 4.06858
H 1.83616 1.50988 2.54427
C 1.05506 -2.47489 3.01483
H 1.16643 -3.35797 2.37720
H 1.58513 -2.66909 3.95576
H -0.00831 -2.33594 3.22261
C 3.50051 -1.20005 1.56261
C 3.81113 -2.31818 0.76211
C 4.50929 -0.24594 1.78763
C 5.07803 -2.46760 0.19842
H 3.05001 -3.06432 0.54858
C 5.78312 -0.39959 1.23791
H 4.29688 0.63776 2.38412
C 6.06661 -1.50754 0.43642
H 5.29465 -3.32960 -0.42691
H 6.54784 0.34985 1.42363
H 7.05430 -1.62394 -0.00197

Cl 4.71333 4.21562 -1.02397

TS-2 (R=Cl)

Rh -3.71007 0.38278 -0.51247
Rh -1.33811 0.03063 0.19067
O -3.44087 2.39153 -0.06168
O -1.28635 2.06206 0.62695
O -4.24390 0.02283 1.46324
O -2.07771 -0.30723 2.10779
O -1.57569 -1.97734 -0.33567
O -3.76323 -1.67152 -0.91759
O -0.82406 0.42349 -1.79518
O -2.99207 0.72454 -2.43942
C 0.64714 -0.25160 0.77912
C 1.45165 0.98300 0.74200
C 1.64456 1.71987 -0.44344
C 1.93005 1.53584 1.94843
C 2.32176 2.93115 -0.43709
H 1.24574 1.33437 -1.37302
C 2.57798 2.76865 1.97156
H 1.76665 1.00814 2.88175
C 2.78447 3.45099 0.77442
H 2.48377 3.47901 -1.35918
H 2.92795 3.19060 2.90753
C 1.10247 -1.32964 1.68656
C 0.20508 -2.28807 2.20089
C 2.47945 -1.48553 1.97613
C 0.66359 -3.33799 2.99335
H -0.84724 -2.20070 1.97836
C 2.93566 -2.54641 2.75216
H 3.19407 -0.78564 1.55837
C 2.02635 -3.47462 3.26902
H -0.04560 -4.06012 3.38814
H 3.99904 -2.65206 2.94930
H 2.37975 -4.30315 3.87721
C -2.32613 2.75723 0.39868
H -2.22709 3.82669 0.63922
C -1.75222 0.67407 -2.63177
H -1.40747 0.86538 -3.65983
C -3.33451 -0.22479 2.29577
H -3.65659 -0.39815 3.33399
C -2.72019 -2.35175 -0.75780
H -2.78877 -3.42083 -1.01396
Si 1.69325 -1.90347 -1.59166
H 0.99236 -0.91325 -0.60057

C 1.43145 -3.60078 -0.83557
H 0.35671 -3.76220 -0.70830
H 1.83594 -4.38695 -1.48498
H 1.89854 -3.68754 0.15000
C 0.87942 -1.70579 -3.27333
H 1.01263 -0.69415 -3.66774
H 1.30464 -2.41914 -3.99020
H -0.19461 -1.89080 -3.18257
C 3.47628 -1.33340 -1.56720
C 3.86940 -0.18435 -2.28036
C 4.44466 -1.98019 -0.77540
C 5.17323 0.30454 -2.20243
H 3.14779 0.34025 -2.90197
C 5.75189 -1.49724 -0.69708
H 4.17458 -2.86679 -0.20793
C 6.11669 -0.35125 -1.40764
H 5.45307 1.19648 -2.75657
H 6.48498 -2.01290 -0.08211
H 7.13248 0.02958 -1.34306
Cl 3.63463 4.98767 0.78653

TS-1 (R=CN)

Rh -3.78567 0.41460 0.29895
Rh -1.35479 0.08459 -0.15954
O -4.16294 -1.01716 -1.15455
O -1.93911 -1.31308 -1.58387
O -3.76361 1.88315 -1.17159
O -1.54119 1.57438 -1.59851
O -0.98892 1.49911 1.33292
O -3.21417 1.82557 1.72678
O -1.38603 -1.37415 1.33256
O -3.61535 -1.08980 1.73359
C 0.66370 -0.19423 -0.61031
C 0.94447 -1.45324 -1.32640
C 0.62906 -2.70832 -0.76488
C 1.40156 -1.41865 -2.66033
C 0.82090 -3.88029 -1.48518
H 0.21884 -2.74502 0.23731
C 1.55541 -2.59593 -3.39282
H 1.60633 -0.46377 -3.13243
C 1.28583 -3.83138 -2.80500
H 0.59061 -4.83720 -1.02436
H 1.89619 -2.54309 -4.42316
H 1.42420 -4.74863 -3.37064
C 1.62022 0.92637 -0.79721

C 1.22389 2.26666 -0.59548
C 2.97741 0.68032 -1.11302
C 2.13090 3.31037 -0.72613
H 0.19929 2.48567 -0.34142
C 3.89283 1.71653 -1.23837
H 3.32629 -0.33730 -1.23123
C 3.47288 3.04229 -1.04706
H 1.80905 4.33583 -0.57628
H 4.93249 1.50204 -1.46013
C -3.17994 -1.53477 -1.74867
H -3.41608 -2.28577 -2.51772
C -2.49280 -1.62144 1.91675
H -2.44253 -2.40769 2.68634
C -2.68227 2.11424 -1.77042
H -2.71135 2.88908 -2.55141
C -1.98965 2.03238 1.91597
H -1.73158 2.77241 2.68964
Si 1.84015 -0.81719 2.19673
H 1.03131 -0.60495 0.87348
C 1.57222 0.75882 3.17735
H 1.81391 1.63526 2.56716
H 0.51895 0.84077 3.45885
H 2.18837 0.77947 4.08459
C 1.20562 -2.37928 3.02792
H 1.33460 -3.25389 2.38167
H 1.75788 -2.56295 3.95788
H 0.14175 -2.27969 3.25621
C 3.59710 -1.01021 1.58215
C 3.95234 -2.12475 0.79495
C 4.56587 -0.01183 1.78949
C 5.22187 -2.22805 0.22637
H 3.22279 -2.90629 0.59624
C 5.84241 -0.11756 1.23376
H 4.32004 0.86850 2.37844
C 6.16933 -1.22273 0.44495
H 5.47290 -3.08903 -0.38747
H 6.57579 0.66573 1.40617
H 7.15885 -1.30232 0.00256
C 4.41636 4.11550 -1.16056
N 5.18650 4.98295 -1.24766

TS-2 (R=CN)

Rh -3.66398 0.35024 -0.55560
Rh -1.30454 0.04792 0.20460
O -3.45393 2.35803 -0.07226

O -1.31579 2.07052 0.68051
O -4.23999 -0.05204 1.40024
O -2.08425 -0.35233 2.09196
O -1.47759 -1.95402 -0.36624
O -3.65884 -1.69623 -0.99174
O -0.74972 0.49917 -1.75694
O -2.90863 0.73441 -2.45684
C 0.65086 -0.19361 0.84815
C 1.43486 1.05189 0.82433
C 1.69762 1.76387 -0.36291
C 1.82319 1.63062 2.05267
C 2.35784 2.98135 -0.33220
H 1.36885 1.35176 -1.30716
C 2.45314 2.86825 2.09238
H 1.60490 1.11195 2.97976
C 2.74102 3.54725 0.89777
H 2.57627 3.50788 -1.25561
H 2.73675 3.30673 3.04368
C 1.12262 -1.27752 1.72383
C 0.24753 -2.28464 2.18580
C 2.49789 -1.38870 2.04783
C 0.72456 -3.33788 2.96116
H -0.80076 -2.22917 1.93581
C 2.97220 -2.45270 2.80689
H 3.19587 -0.65088 1.66875
C 2.08404 -3.42878 3.27146
H 0.03441 -4.09789 3.31615
H 4.03244 -2.52470 3.03306
H 2.45249 -4.25995 3.86684
C -2.36448 2.74485 0.42788
H -2.29848 3.81269 0.68522
C -1.66271 0.72863 -2.61595
H -1.29719 0.94502 -3.63146
C -3.34743 -0.29861 2.25055
H -3.69065 -0.49961 3.27671
C -2.60193 -2.35166 -0.82067
H -2.63563 -3.41722 -1.09667
Si 1.79062 -1.82103 -1.59880
H 1.06998 -0.87067 -0.61863
C 1.61073 -3.53235 -0.84675
H 0.54474 -3.74641 -0.72141
H 2.05297 -4.29834 -1.49529
H 2.07916 -3.60051 0.13996
C 0.95889 -1.66411 -3.27743
H 1.05038 -0.64941 -3.67632

H 1.39975 -2.36306 -3.99853
H -0.10769 -1.88674 -3.17652
C 3.55261 -1.18060 -1.59172
C 3.91518 -0.07028 -2.37866
C 4.52716 -1.72142 -0.73093
C 5.19391 0.48241 -2.30644
H 3.18979 0.37260 -3.05770
C 5.80868 -1.17285 -0.65541
H 4.28318 -2.57785 -0.10773
C 6.14283 -0.06719 -1.44077
H 5.45043 1.34102 -2.92133
H 6.54639 -1.60743 0.01433
H 7.13847 0.36395 -1.37995
C 3.42273 4.80622 0.92964
N 3.98099 5.82695 0.95282

TS-1 (R=F)

Rh 3.73687 -0.30999 0.25442
Rh 1.28553 -0.04762 -0.14191
O 4.03234 1.19533 -1.14120
O 1.78992 1.44601 -1.49913
O 3.73063 -1.71136 -1.28410
O 1.49473 -1.44047 -1.66934
O 1.00005 -1.55307 1.27563
O 3.24107 -1.80204 1.63197
O 1.28587 1.32090 1.43889
O 3.53829 1.12084 1.76041
C -0.77012 0.18531 -0.52987
C -1.07303 1.52117 -1.09232
C -0.91536 2.70437 -0.34453
C -1.37198 1.64271 -2.46488
C -1.10561 3.95178 -0.92825
H -0.61274 2.63470 0.69292
C -1.53402 2.89687 -3.05315
H -1.44719 0.74953 -3.07583
C -1.41996 4.05695 -2.28735
H -0.99432 4.84820 -0.32368
H -1.75613 2.96245 -4.11492
H -1.56191 5.03260 -2.74389
C -1.65084 -0.94221 -0.90820
C -1.22776 -2.28077 -0.73772
C -2.96874 -0.72639 -1.37878
C -2.06052 -3.34722 -1.05563
H -0.23962 -2.47880 -0.35181
C -3.80918 -1.78386 -1.70563

H -3.34960 0.28355 -1.46461
C -3.33795 -3.08183 -1.54204
H -1.73753 -4.37606 -0.93661
H -4.82236 -1.61744 -2.05385
C 3.01829 1.71245 -1.68271
H 3.21317 2.50827 -2.41781
C 2.39737 1.58733 2.00354
H 2.33627 2.32763 2.81706
C 2.64724 -1.93725 -1.88299
H 2.68825 -2.66647 -2.70666
C 2.02934 -2.06927 1.82311
H 1.81184 -2.85467 2.56435
Si -1.99055 0.33362 2.22912
H -1.09746 0.36337 0.91880
C -1.64862 -1.34459 2.99058
H -0.61566 -1.38837 3.34392
H -2.32586 -1.54280 3.83069
H -1.77285 -2.13424 2.24285
C -1.46431 1.79685 3.28364
H -1.76624 2.74655 2.83017
H -1.92680 1.73317 4.27650
H -0.37666 1.79964 3.39314
C -3.72941 0.48380 1.56275
C -4.15818 1.67363 0.94010
C -4.61197 -0.61138 1.57704
C -5.41726 1.75944 0.34626
H -3.49441 2.53313 0.89144
C -5.87760 -0.52543 0.99525
H -4.30403 -1.54969 2.03154
C -6.27956 0.65894 0.37421
H -5.72606 2.68162 -0.13872
H -6.54356 -1.38398 1.01599
H -7.26128 0.72548 -0.08745
F -4.14695 -4.11259 -1.84706

TS-2 (R=F)

Rh -3.64522 0.21224 -0.51415
Rh -1.25541 0.00515 0.19359
O -3.60357 2.12900 0.28123
O -1.41537 1.94713 0.91621
O -4.15850 -0.54252 1.35478
O -1.97747 -0.67823 2.02013
O -1.28060 -1.92379 -0.61052
O -3.46279 -1.73703 -1.26186
O -0.76486 0.73187 -1.70512

O -2.94845 0.96394 -2.32678
C 0.74652 -0.15373 0.78791
C 1.42353 1.15066 0.91342
C 1.57071 2.02979 -0.17897
C 1.80957 1.61132 2.19096
C 2.11570 3.29594 -0.01683
H 1.24114 1.70972 -1.15850
C 2.32202 2.89387 2.37127
H 1.67612 0.96885 3.05431
C 2.48086 3.71302 1.26067
H 2.25184 3.96825 -0.85740
H 2.60364 3.25969 3.35291
C 1.29129 -1.27847 1.58812
C 0.49181 -2.38237 1.94805
C 2.66702 -1.32059 1.91735
C 1.03862 -3.46306 2.63649
H -0.55511 -2.38546 1.68688
C 3.21371 -2.40959 2.58958
H 3.31204 -0.50466 1.61187
C 2.39787 -3.48378 2.95804
H 0.40200 -4.29881 2.91324
H 4.27527 -2.42254 2.82162
H 2.82147 -4.33434 3.48564
C -2.53103 2.54440 0.79617
H -2.55176 3.56769 1.20117
C -1.70744 1.04594 -2.50279
H -1.37669 1.43687 -3.47771
C -3.23672 -0.79106 2.17322
H -3.54802 -1.16019 3.16251
C -2.36446 -2.33392 -1.14554
H -2.31546 -3.35364 -1.55934
Si 1.95745 -1.41916 -1.71283
H 1.15724 -0.61605 -0.62269
C 1.87095 -3.20769 -1.15347
H 0.81822 -3.48545 -1.04534
H 2.34068 -3.86984 -1.89138
H 2.35708 -3.36306 -0.18580
C 1.11398 -1.11501 -3.36284
H 1.14313 -0.05796 -3.64204
H 1.59773 -1.70092 -4.15427
H 0.06396 -1.41209 -3.29103
C 3.67212 -0.67725 -1.60674
C 3.95462 0.56245 -2.21219
C 4.69305 -1.29178 -0.85601
C 5.20353 1.16715 -2.07278

H 3.18879 1.06608 -2.79735
C 5.94552 -0.69196 -0.71593
H 4.50750 -2.24567 -0.36959
C 6.20115 0.54044 -1.32197
H 5.39865 2.12599 -2.54554
H 6.72084 -1.18408 -0.13439
H 7.17437 1.01100 -1.21010
F 3.00053 4.94613 1.42234

TS-1 (R=I)

Rh -4.08969 -1.32670 -0.27159
Rh -1.86270 -0.27437 0.14209
O -4.78298 -0.29104 1.38913
O -2.73925 0.65056 1.78612
O -3.47325 -2.87540 0.97068
O -1.43592 -1.91239 1.34833
O -1.19078 -1.28507 -1.55993
O -3.21734 -2.27670 -1.91460
O -2.48631 1.27603 -1.10659
O -4.52696 0.32150 -1.47673
C -0.00036 0.61312 0.53069
C -0.10359 1.79395 1.41414
C -0.83066 2.94238 1.04048
C 0.39972 1.72703 2.72941
C -1.00037 3.99988 1.92556
H -1.27687 2.98178 0.05350
C 0.19866 2.77732 3.62524
H 0.92391 0.83570 3.05713
C -0.48541 3.92461 3.22500
H -1.55030 4.88206 1.60826
H 0.58713 2.69706 4.63690
C 1.27680 -0.14153 0.54510
C 1.33453 -1.49491 0.15147
C 2.49426 0.49693 0.87333
C 2.54393 -2.18370 0.10549
H 0.42595 -2.01068 -0.11889
C 3.71015 -0.17819 0.82547
H 2.49668 1.54791 1.13627
C 3.72511 -1.51802 0.44004
H 2.56110 -3.22593 -0.19401
H 4.62828 0.34726 1.06099
C -3.96918 0.43446 2.02139
H -4.36933 0.96032 2.90173
C -3.65549 1.21541 -1.61198
H -3.92435 2.06799 -2.25555

C -2.32711 -2.81147 1.48582
H -2.04252 -3.64385 2.14743
C -2.01160 -2.04136 -2.17633
H -1.59451 -2.55129 -3.05916
Si 0.72046 1.85065 -2.15346
H 0.06224 1.25890 -0.84464
C 1.07895 0.34208 -3.20654
H 0.14161 -0.14654 -3.48376
H 1.62938 0.61207 -4.11619
H 1.67101 -0.38482 -2.64059
C -0.50079 3.06988 -2.89631
H -0.08207 3.51641 -3.80680
H -1.43630 2.56171 -3.14173
H -0.72820 3.87988 -2.19541
C 2.27408 2.67582 -1.52079
C 2.19438 3.79274 -0.66470
C 3.54709 2.14519 -1.79906
C 3.34227 4.34814 -0.09955
H 1.22664 4.21771 -0.41164
C 4.69970 2.70722 -1.24786
H 3.64238 1.27098 -2.43784
C 4.59760 3.80544 -0.39112
H 3.25906 5.20081 0.56900
H 5.67305 2.28101 -1.47601
H 5.49247 4.23783 0.04877
H -0.62676 4.74867 3.91887
I 5.59637 -2.55695 0.33173

TS-2 (R=I)

Rh 3.79873 -1.59503 -0.26927
Rh 1.71720 -0.28520 0.16279
O 2.67664 -3.31494 0.04511
O 0.77709 -2.11481 0.46334
O 4.17730 -1.47784 1.77115
O 2.26436 -0.29371 2.16803
O 2.84370 1.43412 -0.21061
O 4.75852 0.23682 -0.55396
O 1.35038 -0.39872 -1.88805
O 3.24699 -1.60631 -2.28461
C -0.04479 0.78214 0.56350
C -1.25535 -0.05546 0.43067
C -1.60704 -0.66745 -0.78857
C -2.00542 -0.39260 1.57387
C -2.69331 -1.52818 -0.88080
H -1.00409 -0.47444 -1.66827

C -3.07869 -1.28104 1.50045
H -1.73274 0.02384 2.53744
C -3.42949 -1.82814 0.26908
H -2.95291 -1.97174 -1.83591
H -3.63421 -1.52941 2.39805
C -0.11035 1.94260 1.48058
C 1.06028 2.48596 2.05451
C -1.34062 2.58788 1.75331
C 0.99508 3.59954 2.88722
H 2.01361 2.02579 1.84771
C -1.40006 3.70632 2.57937
H -2.24976 2.22575 1.28835
C -0.23245 4.21314 3.15381
H 1.90709 3.99275 3.32782
H -2.35612 4.18990 2.75695
H -0.27662 5.08652 3.79937
C 1.45745 -3.18117 0.33518
H 0.89215 -4.11075 0.50184
C 2.18449 -1.02750 -2.61984
H 1.92783 -1.05585 -3.69066
C 3.34405 -0.88042 2.50047
H 3.56997 -0.85276 3.57741
C 4.08378 1.29337 -0.47105
H 4.62725 2.23507 -0.64767
Si -0.28249 2.62845 -1.85837
H -0.02208 1.48952 -0.80573
C 1.09244 3.86259 -1.53919
H 2.05827 3.39925 -1.75672
H 0.97693 4.76356 -2.15406
H 1.10185 4.15531 -0.48385
C -0.26378 1.84040 -3.56448
H -0.43402 2.60260 -4.33503
H 0.69870 1.35566 -3.74531
H -1.04871 1.08266 -3.65933
C -1.96879 3.28245 -1.38126
C -3.11400 2.46886 -1.49490
C -2.11909 4.55631 -0.80315
C -4.35606 2.90521 -1.03400
H -3.03520 1.47302 -1.92284
C -3.36302 5.00383 -0.35535
H -1.25345 5.20276 -0.68318
C -4.48192 4.17515 -0.46271
H -5.22361 2.25624 -1.11789
H -3.45654 5.99218 0.08711
H -5.44925 4.51642 -0.10347

I -5.10917 -3.15399 0.13384

TS-1 (R=Me)

Rh -3.74813 0.28567 0.25942
Rh -1.29244 0.05363 -0.14038
O -4.02927 -1.18506 -1.17553
O -1.78482 -1.39992 -1.54518
O -3.75008 1.72938 -1.23881
O -1.50928 1.49494 -1.62146
O -1.02226 1.51103 1.32963
O -3.26664 1.74656 1.67481
O -1.28619 -1.36834 1.39050
O -3.53858 -1.18711 1.72484
C 0.76813 -0.15054 -0.52888
C 1.07949 -1.45973 -1.14744
C 0.91820 -2.67534 -0.45437
C 1.39290 -1.51925 -2.52051
C 1.11610 -3.89513 -1.09141
H 0.60778 -2.65184 0.58323
C 1.56338 -2.74555 -3.16272
H 1.47510 -0.59893 -3.08860
C 1.44324 -3.93892 -2.45091
H 1.00109 -4.81803 -0.52871
H 1.79732 -2.76320 -4.22388
H 1.59120 -4.89306 -2.94922
C 1.64462 0.99720 -0.85643
C 1.21514 2.32671 -0.64804
C 2.97063 0.80955 -1.31103
C 2.05511 3.40067 -0.91737
H 0.22099 2.50900 -0.26955
C 3.80264 1.89007 -1.58054
H 3.36151 -0.19465 -1.42368
C 3.35960 3.20613 -1.39682
H 1.69317 4.41306 -0.75264
H 4.82047 1.70597 -1.91494
C -3.01043 -1.67434 -1.73393
H -3.19818 -2.45048 -2.49174
C -2.39517 -1.65729 1.94870
H -2.32885 -2.42421 2.73692
C -2.66582 1.98555 -1.82448
H -2.71120 2.74039 -2.62467
C -2.05665 2.00994 1.88258
H -1.84668 2.77522 2.64688
Si 1.98657 -0.39082 2.20987
H 1.08701 -0.39076 0.90172

C 1.70689 1.30517 2.95642
H 0.66997 1.39724 3.28769
H 2.37650 1.47935 3.80787
H 1.88111 2.08244 2.20518
C 1.40228 -1.82835 3.26902
H 1.63783 -2.79017 2.80174
H 1.89159 -1.79908 4.25062
H 0.31925 -1.77091 3.40419
C 3.71991 -0.60864 1.54856
C 4.11084 -1.81501 0.93327
C 4.63978 0.45560 1.56165
C 5.36867 -1.94672 0.34534
H 3.41942 -2.65225 0.88599
C 5.90456 0.32311 0.98719
H 4.36137 1.40709 2.00740
C 6.26851 -0.87690 0.37301
H 5.64684 -2.88077 -0.13536
H 6.59956 1.15845 1.00774
H 7.24931 -0.97932 -0.08411
C 4.25101 4.38011 -1.71678
H 4.09057 4.71967 -2.74895
H 4.04474 5.23245 -1.06034
H 5.31050 4.11977 -1.62063

TS-2 (R=Me)

Rh -3.65570 0.22153 -0.50499
Rh -1.26200 0.00937 0.19214
O -3.60325 2.14283 0.27932
O -1.41198 1.95620 0.90313
O -4.16129 -0.52104 1.37162
O -1.97744 -0.65750 2.02797
O -1.29966 -1.92543 -0.59972
O -3.48388 -1.73337 -1.24278
O -0.77921 0.71723 -1.71520
O -2.96508 0.96093 -2.32544
C 0.74467 -0.15279 0.77646
C 1.42775 1.14969 0.87822
C 1.59461 2.00603 -0.22849
C 1.80455 1.63938 2.14550
C 2.14742 3.26881 -0.07775
H 1.27274 1.66945 -1.20549
C 2.32452 2.92406 2.29054
H 1.65613 1.02015 3.02394
C 2.52320 3.75663 1.18434
H 2.28316 3.89844 -0.95464

H 2.58794 3.28017 3.28378
C 1.28349 -1.26854 1.59530
C 0.48200 -2.36903 1.96082
C 2.65605 -1.30574 1.93667
C 1.02371 -3.44200 2.66522
H -0.56245 -2.37669 1.68996
C 3.19781 -2.38686 2.62580
H 3.30194 -0.49129 1.62948
C 2.38018 -3.45799 2.99866
H 0.38538 -4.27544 2.94518
H 4.25733 -2.39589 2.86739
H 2.80003 -4.30247 3.53898
C -2.52564 2.55647 0.78580
H -2.54024 3.58222 1.18518
C -1.72413 1.03297 -2.50858
H -1.39645 1.41591 -3.48784
C -3.23596 -0.76603 2.18737
H -3.54381 -1.12752 3.18068
C -2.38745 -2.33389 -1.12766
H -2.34451 -3.35625 -1.53588
Si 1.94234 -1.45518 -1.69705
H 1.14209 -0.63339 -0.61789
C 1.82595 -3.23867 -1.12754
H 0.76892 -3.50127 -1.02468
H 2.29238 -3.91173 -1.85770
H 2.30317 -3.39358 -0.15535
C 1.11118 -1.14310 -3.35168
H 1.15427 -0.08646 -3.63076
H 1.59101 -1.73542 -4.14072
H 0.05722 -1.42659 -3.28346
C 3.66950 -0.74377 -1.58565
C 3.98065 0.48465 -2.19973
C 4.67422 -1.37399 -0.82611
C 5.24209 1.06284 -2.06112
H 3.22751 1.00071 -2.79005
C 5.93896 -0.80048 -0.68645
H 4.46604 -2.31891 -0.33164
C 6.22349 0.42072 -1.30207
H 5.45927 2.01344 -2.54097
H 6.70117 -1.30438 -0.09769
H 7.20665 0.87054 -1.19107
C 3.13708 5.12775 1.32898
H 4.18737 5.12596 1.00725
H 2.61319 5.86685 0.71164
H 3.11148 5.47395 2.36729

TS-1 (R=NMe2)

Rh 3.86914 0.45365 -0.28972
Rh 1.43032 0.15007 0.16302
O 4.22164 -0.48642 1.52737
O 1.98982 -0.74277 1.95169
O 3.72241 2.27785 0.70099
O 1.49558 1.99324 1.12876
O 1.06367 1.04921 -1.69551
O 3.29144 1.36662 -2.08779
O 1.58473 -1.68103 -0.82979
O 3.81387 -1.40507 -1.23836
C -0.64577 -0.14042 0.52834
C -0.86177 -1.26722 1.46858
C -0.63870 -2.60958 1.11375
C -1.16199 -0.98219 2.81622
C -0.74521 -3.62815 2.05491
H -0.35545 -2.84653 0.09604
C -1.24836 -2.00266 3.76137
H -1.30119 0.04963 3.12147
C -1.05267 -3.33263 3.38579
H -0.58353 -4.65849 1.74841
H -1.47244 -1.75412 4.79558
H -1.13301 -4.12883 4.12097
C -1.60837 0.97861 0.57015
C -1.28944 2.25919 0.07036
C -2.94289 0.78502 0.99466
C -2.22191 3.28164 0.02122
H -0.29104 2.44796 -0.29548
C -3.89261 1.79245 0.94112
H -3.24626 -0.19290 1.35244
C -3.55591 3.08399 0.45973
H -1.91094 4.24240 -0.37086
H -4.89987 1.57368 1.27519
C 3.22777 -0.85535 2.21082
H 3.45527 -1.34597 3.16980
C 2.72180 -2.02506 -1.28747
H 2.73857 -3.00088 -1.79878
C 2.60624 2.61071 1.17899
H 2.57537 3.57200 1.71555
C 2.07146 1.44431 -2.37168
H 1.82314 1.91527 -3.33662
Si -1.61932 -1.07264 -2.06342
H -0.79828 -0.68908 -0.70465
C -2.17183 0.59831 -2.70274

H -1.29575 1.24840 -2.78069
H -2.62555 0.48724 -3.69598
H -2.89250 1.08393 -2.03921
C -0.39143 -1.94199 -3.18271
H -0.02281 -2.87024 -2.73907
H -0.85515 -2.16951 -4.15120
H 0.46861 -1.28644 -3.34232
C -3.00808 -2.15931 -1.46105
C -2.84212 -3.55316 -1.34593
C -4.23367 -1.60305 -1.04449
C -3.85650 -4.36055 -0.83320
H -1.90839 -4.01612 -1.65567
C -5.24990 -2.40774 -0.52861
H -4.39315 -0.53070 -1.11092
C -5.06187 -3.78753 -0.42020
H -3.70653 -5.43364 -0.75075
H -6.18729 -1.95905 -0.21012
H -5.85153 -4.41455 -0.01475
N -4.48573 4.09845 0.41398
C -4.13332 5.38121 -0.17438
H -3.29437 5.84635 0.35844
H -3.85464 5.28707 -1.23424
H -4.98779 6.05629 -0.10543
C -5.86900 3.83400 0.77527
H -6.32933 3.07643 0.12355
H -5.95097 3.48567 1.81288
H -6.44740 4.75489 0.68633

TS-2 (R=NMe2)

Rh -3.73489 0.70346 -0.47286
Rh -1.40409 0.06189 0.17394
O -3.21978 2.66363 -0.00417
O -1.09765 2.06925 0.61084
O -4.26795 0.38840 1.51181
O -2.13650 -0.17585 2.11002
O -1.89122 -1.90609 -0.34596
O -4.03572 -1.32946 -0.88648
O -0.89040 0.38304 -1.82798
O -3.01282 0.98533 -2.41167
C 0.58735 -0.47539 0.69425
C 1.51371 0.66500 0.63854
C 1.73945 1.39285 -0.54775
C 2.09905 1.18168 1.81338
C 2.52534 2.53003 -0.57612
H 1.27101 1.05496 -1.46406

C 2.85715 2.34416 1.80760
H 1.91922 0.68583 2.76139
C 3.11486 3.04424 0.60622
H 2.67292 3.03000 -1.52593
H 3.24698 2.70592 2.75125
C 0.90373 -1.55896 1.67904
C -0.10082 -2.37804 2.22632
C 2.24716 -1.86715 1.98849
C 0.22354 -3.44110 3.06842
H -1.13458 -2.17969 1.98845
C 2.57091 -2.93831 2.81825
H 3.04185 -1.27532 1.54845
C 1.55710 -3.72896 3.36626
H -0.57108 -4.05442 3.48479
H 3.61386 -3.15806 3.03195
H 1.80552 -4.56518 4.01474
C -2.05365 2.88442 0.42322
H -1.82050 3.93184 0.67106
C -1.79614 0.76939 -2.63569
H -1.45487 0.92770 -3.67122
C -3.37103 0.04321 2.32412
H -3.68841 -0.09280 3.36978
C -3.08233 -2.13424 -0.74379
H -3.29007 -3.18676 -0.99534
Si 1.31937 -2.26218 -1.53288
H 0.74228 -1.12753 -0.51757
C 0.88215 -3.88510 -0.70592
H -0.19099 -3.88266 -0.49382
H 1.11269 -4.72687 -1.37136
H 1.41099 -4.02855 0.24023
C 0.44869 -2.00729 -3.17455
H 0.67852 -1.03208 -3.61113
H 0.74927 -2.79191 -3.88092
H -0.63195 -2.05966 -3.01950
C 3.14479 -1.87999 -1.57686
C 3.63471 -0.80391 -2.34211
C 4.06356 -2.60582 -0.79357
C 4.98653 -0.46289 -2.32467
H 2.95324 -0.21888 -2.95434
C 5.41784 -2.27070 -0.77726
H 3.71779 -3.43561 -0.18346
C 5.88042 -1.19590 -1.54054
H 5.34205 0.37547 -2.91768
H 6.11122 -2.84510 -0.16847
H 6.93413 -0.93013 -1.52398

N 3.90932 4.17744 0.58216
C 4.35351 4.76491 1.83516
H 3.51455 5.09281 2.46895
H 4.95523 4.05098 2.41044
H 4.98283 5.63155 1.62375
C 3.96384 4.98742 -0.62390
H 4.37228 4.41463 -1.46575
H 2.97468 5.36994 -0.91989
H 4.62446 5.83985 -0.45499

TS-1 (R=NO2)

Rh -3.89264 0.50421 0.29590
Rh -1.47446 0.09826 -0.16360
O -4.31707 -0.86816 -1.20184
O -2.10384 -1.21416 -1.64787
O -3.81074 2.01790 -1.12437
O -1.59411 1.65600 -1.53883
O -1.06038 1.44039 1.38332
O -3.27445 1.84955 1.76681
O -1.57798 -1.42309 1.25771
O -3.78993 -1.05340 1.67961
C 0.53684 -0.25809 -0.59924
C 0.77649 -1.48906 -1.37320
C 0.34571 -2.74779 -0.90293
C 1.31527 -1.41409 -2.67476
C 0.49563 -3.88840 -1.68019
H -0.11401 -2.80818 0.07681
C 1.42774 -2.55681 -3.46675
H 1.61642 -0.45270 -3.07685
C 1.03652 -3.79926 -2.96883
H 0.17458 -4.85057 -1.29011
H 1.83130 -2.47273 -4.47196
H 1.14126 -4.69105 -3.58058
C 1.55882 0.81909 -0.70131
C 1.21772 2.17414 -0.50459
C 2.92049 0.49768 -0.91243
C 2.19015 3.16803 -0.53101
H 0.18782 2.44478 -0.33528
C 3.90374 1.47948 -0.92211
H 3.21513 -0.53685 -1.03512
C 3.52105 2.80495 -0.73233
H 1.93526 4.21063 -0.38716
H 4.94845 1.22985 -1.05562
C -3.35114 -1.39205 -1.81792
H -3.61082 -2.10552 -2.61457

C -2.69331 -1.64425 1.83568
H -2.67654 -2.46375 2.57092
C -2.71554 2.23846 -1.70156
H -2.71101 3.04419 -2.45108
C -2.04414 1.99904 1.97248
H -1.76343 2.70598 2.76879
Si 1.66955 -0.99304 2.18761
H 0.85574 -0.75501 0.86961
C 1.75769 0.70066 2.98823
H 2.22718 1.42160 2.31023
H 0.74568 1.05924 3.19450
H 2.33056 0.67745 3.92323
C 0.76455 -2.29778 3.19240
H 0.61491 -3.21142 2.60785
H 1.34256 -2.55552 4.08843
H -0.21784 -1.92658 3.49486
C 3.33172 -1.58263 1.56126
C 3.43289 -2.78168 0.82742
C 4.48955 -0.79485 1.70176
C 4.63870 -3.16789 0.24202
H 2.55699 -3.40887 0.68451
C 5.70202 -1.18432 1.13007
H 4.44311 0.14383 2.24816
C 5.77551 -2.36799 0.39169
H 4.69107 -4.08801 -0.33378
H 6.58375 -0.56049 1.25031
H 6.71495 -2.66648 -0.06592
N 4.55968 3.85292 -0.72804
O 4.19421 5.01760 -0.57815
O 5.72889 3.49609 -0.87105

TS-2 (R=NO2)

Rh -3.72184 0.59011 -0.55732
Rh -1.40355 0.08209 0.20834
O -3.31482 2.58381 -0.14434
O -1.19682 2.11441 0.57661
O -4.32540 0.31580 1.41282
O -2.20064 -0.13656 2.11589
O -1.78947 -1.91800 -0.24930
O -3.92658 -1.45999 -0.91376
O -0.82764 0.36698 -1.78074
O -2.94974 0.83810 -2.47180
C 0.52576 -0.33094 0.84341
C 1.44032 0.81672 0.73230
C 1.75673 1.42255 -0.50051

C 1.91697 1.41715 1.91961
C 2.55417 2.55475 -0.55482
H 1.36083 0.99768 -1.41216
C 2.68849 2.57235 1.87691
H 1.66077 0.98159 2.87905
C 3.01281 3.11855 0.63687
H 2.82165 3.01183 -1.49928
H 3.05295 3.04188 2.78216
C 0.90147 -1.41246 1.76407
C -0.06556 -2.28788 2.30536
C 2.26586 -1.65937 2.05944
C 0.31590 -3.34285 3.12961
H -1.10895 -2.12831 2.08037
C 2.64256 -2.72694 2.86629
H 3.02921 -1.02738 1.61993
C 1.66623 -3.56899 3.40984
H -0.44266 -3.99900 3.54664
H 3.69481 -2.90585 3.06890
C -2.17982 2.87972 0.31591
H -2.00164 3.94506 0.52550
C -1.71556 0.67644 -2.64055
H -1.33878 0.81287 -3.66566
C -3.45381 0.03309 2.27326
H -3.80766 -0.08906 3.30809
C -2.94763 -2.21583 -0.69566
H -3.09467 -3.28450 -0.91661
Si 1.43583 -2.21188 -1.53269
H 0.83854 -1.13614 -0.60533
C 1.07993 -3.84871 -0.68316
H -0.00174 -3.94018 -0.54396
H 1.43378 -4.69467 -1.28476
H 1.54612 -3.90643 0.30537
C 0.60468 -2.05620 -3.21182
H 0.80050 -1.08056 -3.66663
H 0.95863 -2.83653 -3.89634
H -0.47843 -2.15626 -3.09227
C 3.25860 -1.77398 -1.57218
C 3.72384 -0.72235 -2.38557
C 4.18689 -2.40755 -0.72386
C 5.05692 -0.31351 -2.34992
H 3.03632 -0.21044 -3.05547
C 5.52302 -2.00433 -0.68594
H 3.86394 -3.22281 -0.08159
C 5.95894 -0.95331 -1.49614
H 5.39204 0.50359 -2.98317

H 6.22364 -2.50907 -0.02564
H 6.99735 -0.63463 -1.46398
H 1.95876 -4.40192 4.04365
N 3.86009 4.31704 0.58198
O 4.23870 4.80076 1.64950
O 4.14472 4.76605 -0.52928

TS-1 (R=OCF3)

Rh 3.93455 -1.29279 0.36485
Rh 1.74115 -0.22164 -0.16209
O 4.76645 -0.13028 -1.13882
O 2.76095 0.84889 -1.62592
O 3.44628 -2.74428 -1.04102
O 1.45005 -1.74953 -1.53689
O 0.91804 -1.39544 1.35725
O 2.92388 -2.37554 1.83916
O 2.22295 1.22268 1.27544
O 4.24087 0.25650 1.72968
C -0.08139 0.68946 -0.65942
C 0.09924 1.92181 -1.45627
C 0.80554 3.03579 -0.95953
C -0.28903 1.93595 -2.81300
C 1.06102 4.13809 -1.76579
H 1.17199 3.01189 0.05841
C 0.00269 3.02904 -3.62853
H -0.79005 1.07188 -3.23558
C 0.66157 4.14215 -3.10688
H 1.59170 4.99207 -1.35306
H -0.29415 3.00902 -4.67368
H 0.87319 5.00010 -3.73909
C -1.34689 -0.05741 -0.84904
C -1.41941 -1.44715 -0.61654
C -2.53216 0.60775 -1.24408
C -2.60376 -2.14707 -0.82576
H -0.54006 -1.97924 -0.28945
C -3.72108 -0.08096 -1.45214
H -2.52061 1.68187 -1.38100
C -3.73503 -1.45867 -1.25122
H -2.65997 -3.21951 -0.67326
H -4.62619 0.43241 -1.75215
C 4.00715 0.64991 -1.77377
H 4.47896 1.24831 -2.56798
C 3.34955 1.12938 1.86827
H 3.55316 1.92559 2.60148
C 2.35413 -2.63889 -1.65601

H 2.13641 -3.41827 -2.40206
C 1.69188 -2.18425 1.99339
H 1.20161 -2.77115 2.78592
Si -0.86803 1.77490 2.13052
H -0.27947 1.24263 0.76437
C -0.27593 0.60588 3.47227
H 0.81607 0.61841 3.51151
H -0.67687 0.90824 4.44761
H -0.58349 -0.42142 3.26179
C -0.26709 3.54366 2.35182
H -0.51274 4.17268 1.49015
H -0.72224 3.99103 3.24408
H 0.82052 3.55079 2.47611
C -2.71196 1.67441 1.84795
C -3.42914 2.75512 1.29848
C -3.41111 0.47865 2.10038
C -4.79025 2.64470 1.00986
H -2.92074 3.69343 1.08799
C -4.76968 0.36274 1.81210
H -2.88603 -0.37953 2.51177
C -5.46026 1.44524 1.26160
H -5.32561 3.48998 0.58511
H -5.28465 -0.57438 1.99202
H -6.51475 1.34668 1.02031
O -4.88792 -2.20337 -1.54472
C -5.90211 -2.17831 -0.64917
F -6.84603 -3.01443 -1.07473
F -5.50495 -2.56376 0.58309
F -6.45025 -0.94874 -0.51833

TS-2 (R=OCF3)

Rh -3.59251 1.28833 -0.56148
Rh -1.48541 0.16807 0.17581
O -2.62720 3.09918 -0.23765
O -0.70910 2.07186 0.46322
O -4.19274 1.26465 1.42993
O -2.26472 0.24346 2.10509
O -2.43121 -1.65711 -0.20292
O -4.37507 -0.63344 -0.83075
O -0.90372 0.21013 -1.83036
O -2.82806 1.23605 -2.49908
C 0.27992 -0.74207 0.80066
C 1.46202 0.12662 0.65966
C 1.88301 0.63218 -0.58639
C 2.10407 0.59876 1.82464

C 2.92225 1.54773 -0.67321
H 1.37298 0.30913 -1.48399
C 3.11906 1.54917 1.74742
H 1.77833 0.24492 2.79652
C 3.52474 2.00280 0.49832
H 3.26203 1.92688 -1.63029
H 3.60145 1.93457 2.63901
C 0.35692 -1.84006 1.78502
C -0.80582 -2.39417 2.36047
C 1.60375 -2.43296 2.10023
C -0.71956 -3.47388 3.23592
H -1.76849 -1.97037 2.11903
C 1.68337 -3.52357 2.96002
H 2.50684 -2.05163 1.63729
C 0.52009 -4.04426 3.53612
H -1.62548 -3.87989 3.67707
H 2.64984 -3.97010 3.17731
H 0.58010 -4.89474 4.21013
C -1.44298 3.07495 0.19256
H -0.96564 4.05234 0.35977
C -1.69569 0.72630 -2.68456
H -1.32805 0.71347 -3.72230
C -3.41070 0.77049 2.28136
H -3.75274 0.78714 3.32748
C -3.64028 -1.62777 -0.61063
H -4.08930 -2.61734 -0.79022
Si 0.61310 -2.85092 -1.45155
H 0.34321 -1.60996 -0.55382
C -0.18190 -4.28836 -0.54315
H -1.23990 -4.05605 -0.38701
H -0.10298 -5.21511 -1.12466
H 0.26821 -4.45493 0.44009
C -0.16086 -2.53008 -3.13327
H 0.29165 -1.66672 -3.62973
H -0.04702 -3.40635 -3.78327
H -1.22681 -2.31992 -3.00564
C 2.48379 -2.93555 -1.50201
C 3.21714 -2.07911 -2.34606
C 3.20537 -3.78243 -0.63872
C 4.61171 -2.06392 -2.32761
H 2.69503 -1.41071 -3.02719
C 4.60117 -3.77287 -0.61797
H 2.67353 -4.45449 0.02978
C 5.30663 -2.91062 -1.46048
H 5.15619 -1.39177 -2.98551

H 5.13818 -4.43705 0.05441
H 6.39315 -2.89850 -1.44221
O 4.59816 2.90403 0.42229
C 4.27349 4.21759 0.30837
F 5.40875 4.91546 0.28320
F 3.52166 4.64720 1.33993
F 3.58332 4.47869 -0.82134

TS-1 (R=OMe)

Rh 3.77513 0.05026 -0.34073
Rh 1.32684 0.05469 0.15750
O 4.04089 -0.92749 1.47201
O 1.80014 -0.92468 1.92535
O 3.88128 1.87845 0.64719
O 1.65118 1.85892 1.14430
O 1.04756 1.01926 -1.68030
O 3.28814 1.03330 -2.12692
O 1.23205 -1.76035 -0.87063
O 3.47814 -1.78645 -1.28274
C -0.73303 0.03671 0.59762
C -1.09463 -1.08757 1.48994
C -1.04600 -2.43291 1.08098
C -1.35078 -0.81690 2.85040
C -1.27919 -3.46365 1.98453
H -0.80366 -2.66161 0.05127
C -1.55656 -1.85414 3.75832
H -1.35684 0.21170 3.19588
C -1.53483 -3.18200 3.32971
H -1.25547 -4.49372 1.63841
H -1.73998 -1.61977 4.80358
H -1.71070 -3.98941 4.03531
C -1.56111 1.25573 0.66224
C -1.08611 2.49443 0.18703
C -2.90745 1.20654 1.10665
C -1.88257 3.63511 0.19001
H -0.07849 2.56155 -0.19463
C -3.71453 2.33029 1.10309
H -3.32278 0.26227 1.43953
C -3.20495 3.56188 0.65360
H -1.46946 4.56833 -0.17440
H -4.74487 2.28889 1.44203
C 3.02040 -1.18298 2.16699
H 3.20070 -1.70389 3.11984
C 2.31145 -2.25206 -1.33388
H 2.19416 -3.21676 -1.85244

C 2.82956 2.33941 1.16151
H 2.92989 3.29322 1.70260
C 2.08323 1.27685 -2.38056
H 1.87735 1.78164 -3.33812
Si -1.91319 -0.68899 -2.06477
H -1.05548 -0.49167 -0.74368
C -2.22139 1.05802 -2.67260
H -1.25700 1.56015 -2.79227
H -2.74176 1.04383 -3.63838
H -2.81601 1.64294 -1.96478
C -0.86073 -1.71962 -3.22721
H -0.64850 -2.70837 -2.81107
H -1.36105 -1.84421 -4.19565
H 0.09632 -1.21333 -3.38054
C -3.45634 -1.55373 -1.46209
C -3.49418 -2.95407 -1.31768
C -4.58915 -0.81858 -1.06151
C -4.61518 -3.59579 -0.79226
H -2.63691 -3.55344 -1.61480
C -5.71256 -1.45666 -0.53390
H -4.59361 0.26382 -1.15337
C -5.72583 -2.84644 -0.39617
H -4.62218 -4.67743 -0.68761
H -6.57590 -0.87042 -0.22990
H -6.59853 -3.34435 0.01811
O -4.06825 4.60578 0.69986
C -3.61336 5.88650 0.27731
H -3.32266 5.87984 -0.78120
H -4.45824 6.56275 0.41683
H -2.76620 6.23046 0.88447

TS-2 (R=OMe)

Rh -3.70403 0.41069 -0.51099
Rh -1.32733 0.05373 0.18061
O -3.39721 2.43425 -0.15776
O -1.23186 2.10506 0.50033
O -4.21976 0.15218 1.48570
O -2.04788 -0.15383 2.12480
O -1.60750 -1.98099 -0.22115
O -3.79151 -1.66322 -0.81232
O -0.83193 0.31597 -1.83244
O -2.99934 0.65478 -2.46033
C 0.69340 -0.24162 0.73301
C 1.50767 0.98035 0.60415
C 1.68090 1.64401 -0.63150

C 2.01965 1.61285 1.75300
C 2.36127 2.84337 -0.71616
H 1.26052 1.20362 -1.52664
C 2.67412 2.84277 1.68505
H 1.87489 1.15337 2.72482
C 2.86386 3.45934 0.44321
H 2.51187 3.33926 -1.66998
H 3.03075 3.30091 2.60017
C 1.12861 -1.23895 1.75571
C 0.21538 -2.12671 2.35531
C 2.50006 -1.39428 2.06126
C 0.65314 -3.10816 3.24350
H -0.83525 -2.04259 2.12301
C 2.93700 -2.38538 2.93641
H 3.22717 -0.74832 1.58200
C 2.01178 -3.24515 3.53566
H -0.07108 -3.77668 3.70102
H 3.99832 -2.48980 3.14649
H 2.34896 -4.01934 4.22001
C -2.26556 2.80318 0.25886
H -2.14572 3.88230 0.44144
C -1.76458 0.55646 -2.66631
H -1.43045 0.68944 -3.70741
C -3.30110 -0.05210 2.32079
H -3.61344 -0.16221 3.37087
C -2.76247 -2.35468 -0.61516
H -2.85585 -3.43577 -0.80627
Si 1.60796 -2.06792 -1.44326
H 0.94248 -0.96047 -0.48599
C 1.38324 -3.67852 -0.51228
H 0.31974 -3.79711 -0.28423
H 1.71052 -4.52439 -1.13013
H 1.93279 -3.70080 0.43287
C 0.67564 -2.03031 -3.07063
H 0.77910 -1.06938 -3.58069
H 1.04643 -2.82393 -3.73190
H -0.38772 -2.19424 -2.87747
C 3.37662 -1.48583 -1.57225
C 3.72643 -0.43858 -2.44673
C 4.38542 -2.02430 -0.74972
C 5.02998 0.05382 -2.49772
H 2.97241 0.00417 -3.09295
C 5.69180 -1.53652 -0.80004
H 4.14823 -2.82834 -0.05842
C 6.01483 -0.49432 -1.67235

H 5.27690 0.86619 -3.17583
H 6.45679 -1.96772 -0.15953
H 7.03083 -0.11025 -1.70918
O 3.51040 4.64139 0.24967
C 4.02839 5.31998 1.38424
H 4.49449 6.23014 1.00227
H 3.23215 5.58895 2.09137
H 4.78379 4.71610 1.90528

TS-1 (R=SCF3)

Rh 4.18374 -1.07606 0.31876
Rh 1.90565 -0.17184 -0.16939
O 4.92978 0.33993 -1.00108
O 2.84920 1.17799 -1.43750
O 3.89626 -2.38836 -1.27203
O 1.83332 -1.52802 -1.74483
O 1.16402 -1.59089 1.16758
O 3.24333 -2.42608 1.60757
O 2.18249 1.09890 1.46390
O 4.28854 0.31443 1.86780
C 0.02820 0.59600 -0.64281
C 0.07534 1.98472 -1.13537
C 0.55356 3.04439 -0.33877
C -0.21879 2.24936 -2.48924
C 0.68687 4.32445 -0.86215
H 0.82889 2.84301 0.68942
C -0.04815 3.52807 -3.01905
H -0.55131 1.44082 -3.13188
C 0.38897 4.57425 -2.20663
H 1.03848 5.13076 -0.22380
H -0.26740 3.70515 -4.06841
H 0.50460 5.57402 -2.61600
C -1.12904 -0.23982 -1.04314
C -1.07333 -1.64704 -0.96317
C -2.36874 0.35677 -1.36830
C -2.21401 -2.41887 -1.15531
H -0.14227 -2.12974 -0.71000
C -3.50781 -0.41085 -1.58058
H -2.45280 1.43615 -1.40333
C -3.43870 -1.80065 -1.43750
H -2.16524 -3.49754 -1.04524
H -4.45435 0.07113 -1.79446
C 4.11301 1.12006 -1.55991
H 4.54317 1.86006 -2.25188
C 3.29536 1.05324 2.08354

H 3.38467 1.76036 2.92301
C 2.82570 -2.30603 -1.92631
H 2.71700 -2.99825 -2.77525
C 1.99573 -2.37259 1.73729
H 1.54446 -3.09427 2.43636
Si -1.31893 0.85595 2.13061
H -0.37784 0.80679 0.88124
C -1.78167 -0.92507 2.49888
H -2.35745 -1.36383 1.67773
H -0.87275 -1.51896 2.62251
H -2.38387 -0.99114 3.41361
C -0.32649 1.70202 3.48470
H -0.06680 2.72743 3.20063
H -0.90026 1.74043 4.41878
H 0.60502 1.15617 3.65807
C -2.79024 1.85150 1.54016
C -2.64506 3.20121 1.16201
C -4.04942 1.24766 1.35962
C -3.71168 3.91554 0.61619
H -1.68421 3.69713 1.27095
C -5.12235 1.96166 0.82386
H -4.19699 0.20214 1.61334
C -4.95330 3.29553 0.44673
H -3.57423 4.95190 0.31931
H -6.07928 1.46763 0.68572
H -5.78437 3.85036 0.01899
C -5.75573 -2.16608 -0.05171
F -6.82385 -2.93647 0.19315
F -4.95747 -2.17968 1.03543
F -6.18696 -0.89267 -0.19407
S -4.92266 -2.80474 -1.54500

TS-2 (R=SCF3)

Rh 3.36795 -1.82722 -0.60780
Rh 1.52902 -0.34641 0.20339
O 2.13943 -3.43382 -0.13072
O 0.46225 -2.07609 0.62087
O 4.10089 -1.82288 1.33945
O 2.40537 -0.48602 2.08588
O 2.73847 1.28076 -0.30268
O 4.43303 -0.07300 -1.01721
O 0.81842 -0.36800 -1.76218
O 2.50028 -1.72371 -2.49557
C -0.00605 0.88181 0.86835
C -1.33542 0.25345 0.78800

C -1.92118 -0.14125 -0.43035
C -1.99022 -0.09531 1.98823
C -3.13277 -0.81590 -0.45352
H -1.41440 0.08703 -1.35744
C -3.17730 -0.82267 1.96857
H -1.54306 0.17319 2.93941
C -3.76125 -1.17481 0.74778
H -3.58662 -1.08592 -1.40055
H -3.65229 -1.11329 2.90016
C 0.15648 2.00762 1.80455
C 1.42641 2.37420 2.29916
C -0.95437 2.81391 2.15565
C 1.57122 3.48011 3.13264
H 2.28928 1.78558 2.02761
C -0.80206 3.92747 2.97472
H -1.93211 2.57659 1.75151
C 0.46285 4.26063 3.47107
H 2.55526 3.74081 3.51172
H -1.66575 4.53813 3.22328
H 0.58357 5.12945 4.11272
C 1.00293 -3.19723 0.35927
H 0.38222 -4.07394 0.59778
C 1.45502 -1.04143 -2.63620
H 1.02403 -1.00957 -3.64873
C 3.46491 -1.18200 2.21449
H 3.86970 -1.21683 3.23747
C 3.89397 1.03706 -0.78694
H 4.48786 1.93133 -1.03290
Si -0.01148 2.94628 -1.49132
H 0.06007 1.72633 -0.53993
C 0.95284 4.30555 -0.62480
H 1.97912 3.96064 -0.46440
H 0.97716 5.21772 -1.23349
H 0.52962 4.55281 0.35365
C 0.74204 2.45558 -3.14184
H 0.17541 1.65250 -3.62150
H 0.77775 3.31430 -3.82335
H 1.76090 2.09134 -2.97994
C -1.85132 3.28827 -1.59507
C -2.66380 2.58582 -2.50568
C -2.48088 4.17544 -0.70087
C -4.04790 2.75808 -2.52308
H -2.21208 1.89079 -3.21012
C -3.86565 4.35125 -0.71363
H -1.88386 4.73233 0.01669

C -4.65156 3.64028 -1.62361
H -4.65550 2.20464 -3.23418
H -4.33160 5.04277 -0.01628
H -5.73003 3.77375 -1.63298
C -4.71034 -3.68320 0.17986
F -3.78228 -4.19483 1.00395
F -4.16269 -3.63928 -1.05091
F -5.75755 -4.52373 0.14093
S -5.32311 -2.05445 0.73452

TS-3

Rh -0.14295 0.01475 0.09776
Rh -0.19853 0.31177 -2.52950
P 2.58202 0.86064 -1.18911
P 0.48631 -2.58190 -1.53673
P -2.92088 -0.47371 -1.27909
P -0.72711 2.91429 -0.97317
O 1.77818 0.82420 0.11482
O 1.83929 0.80597 -2.51391
O 3.47869 2.23140 -1.27657
O 3.65829 -0.37436 -1.03067
O 0.62554 -1.93335 -0.15377
O 0.30516 -1.70780 -2.76125
O 1.79293 -3.51014 -1.89474
O -0.74006 -3.66806 -1.34265
O -2.06592 -0.86347 -0.06416
O -2.23989 -0.17511 -2.60001
O -3.82078 0.78058 -0.70928
O -3.99950 -1.66022 -1.64952
O -0.86859 1.97319 0.23134
O -0.77018 2.32470 -2.37234
O -1.92938 4.03799 -0.94258
O 0.65684 3.75412 -0.70777
C 6.35336 1.24944 -1.16270
C 5.51796 1.89732 -0.10321
C 4.22729 2.44594 -0.19566
C 3.74004 3.16147 0.84577
C 4.48608 3.31271 2.03002
C 5.72476 2.67057 2.16742
H 6.28563 2.73954 3.08987
C 6.20959 1.93740 1.08540
C 7.45787 1.13688 1.03373
H 7.49668 0.40576 1.86936
H 8.34074 1.81078 1.06336
C 7.35499 0.41772 -0.32398

H 8.35128 0.31960 -0.80852
H 6.94811 -0.60520 -0.15548
C 5.65170 0.43243 -2.20139
C 4.49255 -0.47519 -2.06575
C 4.14862 -1.25708 -3.18010
H 3.29793 -1.91919 -3.13534
C 4.84331 -1.11791 -4.39257
C 5.86788 -0.17030 -4.53050
H 6.36219 -0.03891 -5.48445
C 6.23917 0.60479 -3.43099
C 7.27248 1.66811 -3.41378
H 7.08644 2.41648 -4.21404
H 8.28102 1.21637 -3.52986
C 7.10700 2.29872 -2.02174
H 8.09050 2.58957 -1.59128
H 6.49091 3.22102 -2.12311
C 0.63605 -6.34485 -1.80545
C 1.49810 -5.61426 -0.81646
C 2.11682 -4.35683 -0.92007
C 3.00351 -3.95568 0.09669
C 3.20508 -4.77009 1.22278
C 2.49900 -5.97292 1.35675
H 2.60475 -6.57489 2.24999
C 1.63416 -6.36277 0.33248
C 0.72308 -7.53420 0.34696
H 1.31116 -8.47198 0.25150
H 0.11050 -7.54993 1.27421
C -0.16222 -7.30569 -0.89036
H -1.10875 -6.81688 -0.56490
H -0.42081 -8.26622 -1.38798
C -0.25994 -5.55138 -2.71249
C -0.97336 -4.37037 -2.44988
C -1.82798 -3.86259 -3.44729
H -2.38546 -2.95369 -3.27338
C -1.91992 -4.49907 -4.69498
C -1.13476 -5.62603 -4.96842
H -1.16431 -6.09177 -5.94504
C -0.29938 -6.12307 -3.96644
C 0.66204 -7.24310 -4.10417
H 1.30636 -7.10736 -4.99951
H 0.11253 -8.20708 -4.16239
C 1.49183 -7.15584 -2.81381
H 2.43498 -6.60729 -3.03744
H 1.76413 -8.16771 -2.44073
C -6.61578 -0.34777 -0.64376

C -6.00381 0.61946 -1.61155	H 1.96971 3.40804 -2.99082
C -4.72945 1.20004 -1.58588	C 1.46572 5.23417 -3.88922
C -4.39363 2.13884 -2.58022	C 0.77846 6.45570 -3.78906
H -3.41794 2.60030 -2.58647	H 0.72159 7.11851 -4.64299
C -5.30788 2.44946 -3.59945	C 0.15747 6.80044 -2.58406
C -6.54891 1.80207 -3.65127	C -0.65014 8.02121 -2.34106
H -7.23975 2.00287 -4.45997	H -1.43497 8.14088 -3.11892
C -6.86688 0.87782 -2.65434	H 0.01154 8.91362 -2.31433
C -8.08439 0.03567 -2.59676	C -1.27716 7.75580 -0.96556
H -8.25677 -0.48213 -3.56498	H -1.31799 8.68389 -0.35390
H -8.96284 0.66049 -2.32748	H -2.32042 7.39560 -1.11486
C -7.76174 -0.97459 -1.48492	H -4.35616 -4.56798 1.72296
H -8.66671 -1.21106 -0.88297	H -3.41116 4.21720 3.32913
H -7.41173 -1.91937 -1.95968	H 4.08197 3.88808 2.85267
C -5.74520 -1.40993 -0.03437	H 3.87288 -4.44726 2.01043
C -4.66462 -2.12215 -0.59153	H -2.56810 -4.09395 -5.46112
C -4.20996 -3.28993 0.05247	H -5.03887 3.16345 -4.36698
H -3.41359 -3.87041 -0.39487	H 1.93671 4.96537 -4.82616
C -4.72619 -3.67625 1.23412	H 4.55001 -1.71418 -5.24683
C -5.74725 -2.83869 1.90414	C -0.10251 -0.08278 2.23649
H -6.11722 -3.09099 2.88911	C 1.28565 -0.09156 2.75809
C -6.20332 -1.71116 1.22750	C -1.16198 0.53567 3.05327
C -7.19938 -0.72306 1.71341	C 2.15569 -1.16819 2.48675
H -6.88756 -0.29256 2.68942	C 1.81493 1.02624 3.44062
H -8.19784 -1.20293 1.79913	C -2.42442 0.84774 2.48716
C -7.19498 0.35059 0.60996	C -1.02378 0.73895 4.44639
H -6.52350 1.18233 0.92259	C 3.46537 -1.17357 2.93716
H -8.21250 0.76768 0.44326	H 1.78873 -1.99284 1.88984
C -0.41286 6.64004 -0.32619	C 3.12100 1.02357 3.91854
C -1.22375 5.75106 0.56703	H 1.20036 1.90597 3.59163
C -2.03453 4.65064 0.23737	C -3.46347 1.33624 3.25733
C -2.85926 4.11653 1.24223	H -2.58505 0.68693 1.43334
C -2.78425 4.62757 2.54956	C -2.05248 1.25610 5.22816
C -1.86000 5.63855 2.85830	H -0.09736 0.46736 4.93898
H -1.76137 5.99079 3.87729	C 3.92900 -0.08455 3.67615
C -1.10303 6.16094 1.87975	H 4.13260 -2.00091 2.73293
C -0.05371 7.20047 2.04207	H 3.51792 1.86213 4.47748
H -0.52325 8.17695 2.28896	C -3.28708 1.56260 4.63265
H 0.67771 6.91060 2.82720	H -4.42618 1.56656 2.81162
C 0.61748 7.24507 0.65801	H -1.89045 1.40547 6.28964
H 1.53205 6.60998 0.68550	H -3.34913 -3.22106 3.92105
H 0.92265 8.27967 0.38695	C -2.76446 -2.33429 3.64582
C 0.23266 6.01522 -1.52147	Si -1.03315 -2.85198 3.13039
C 0.87873 4.68946 -1.63209	H -2.76906 -1.64542 4.49441
C 1.50135 4.37076 -2.85438	H -3.26434 -1.82702 2.81684

C 0.07978 -3.15681 4.61390
C -1.02257 -4.28520 1.91957
H -0.45707 -1.56618 2.36201
C 1.20711 -3.99517 4.51416
C -0.17499 -2.55790 5.86408
H -1.32304 -3.96338 0.92051
H -0.02409 -4.71954 1.83481
H -1.70225 -5.07425 2.26582
C 2.04504 -4.22485 5.60484
H 1.43406 -4.48244 3.57133
C 0.65967 -2.78314 6.95964
H -1.03829 -1.91023 5.98881
H 2.90834 -4.87612 5.49776
C 1.77381 -3.61504 6.83141
H 0.44031 -2.31134 7.91403
H 2.42608 -3.78866 7.68307
H -3.51970 3.28835 1.02022
H 2.75737 3.60836 0.78635
H 3.50239 -2.99704 0.03383
O -4.35673 2.06751 5.29201
C -4.23327 2.35768 6.67888
H -5.19612 2.77107 6.98246
H -4.02684 1.45113 7.26159
H -3.44410 3.09721 6.86292
N 5.29173 -0.11153 4.21112
O 6.01046 -1.06461 3.90420
O 5.64063 0.81578 4.94473

TS-4

Rh 0.11885 -0.00436 0.08427
Rh 0.28878 0.44131 -2.51873
P 2.22035 2.14847 -0.81699
P 2.15183 -1.82098 -1.47418
P -1.88843 -1.58746 -1.65841
P -1.76801 2.25364 -1.00985
O 1.31904 1.67121 0.32279
O 1.77557 1.92097 -2.24924
O 2.48419 3.76041 -0.66831
O 3.67962 1.44742 -0.49135
O 1.79363 -1.27406 -0.08486
O 1.75898 -1.04191 -2.71150
O 3.77687 -2.01527 -1.62637
O 1.53699 -3.34971 -1.44302
O -1.16399 -1.62014 -0.30254
O -1.14164 -1.05901 -2.87177

O -3.28031 -0.77174 -1.38530
O -2.36409 -3.11240 -2.06607
O -1.53486 1.28790 0.15778
O -1.14998 1.94089 -2.36452
O -3.36821 2.40935 -1.34358
O -1.28286 3.71779 -0.43587
C 5.40468 4.01124 0.10771
C 4.17591 4.12990 0.95626
C 2.82347 4.11079 0.57276
C 1.86924 4.34126 1.50782
C 2.21335 4.53958 2.85870
C 3.55392 4.46198 3.26055
H 3.82563 4.55742 4.30341
C 4.52213 4.22640 2.28489
C 5.97012 3.98773 2.50434
H 6.13338 3.20546 3.27627
H 6.46913 4.93580 2.79830
C 6.45805 3.50752 1.12533
H 7.48454 3.87553 0.90674
H 6.48120 2.39398 1.12452
C 5.32714 3.14827 -1.11451
C 4.64648 1.85001 -1.31514
C 4.87492 1.17605 -2.52514
H 4.38240 0.23547 -2.72317
C 5.67280 1.75571 -3.52473
C 6.22697 3.03279 -3.35367
H 6.80238 3.48653 -4.15034
C 6.02456 3.71542 -2.15307
C 6.51052 5.07726 -1.82439
H 6.20111 5.80826 -2.60220
H 7.61653 5.06753 -1.71777
C 5.82921 5.38241 -0.48057
H 6.50415 5.95541 0.19268
H 4.92879 6.00804 -0.67606
C 4.03244 -5.07188 -1.67718
C 4.31191 -4.08840 -0.57778
C 4.32264 -2.68383 -0.61286
C 4.82209 -1.98610 0.50222
C 5.22568 -2.68465 1.65141
C 5.09068 -4.07831 1.71202
H 5.34750 -4.61633 2.61526
C 4.61822 -4.75717 0.58739
C 4.37389 -6.21679 0.47405
H 5.34216 -6.76168 0.46942
H 3.72821 -6.58177 1.30091

C 3.65994 -6.34789 -0.88214
H 2.56078 -6.37399 -0.70378
H 3.94920 -7.28631 -1.40442
C 2.99973 -4.72721 -2.70989
C 1.79340 -4.02590 -2.56056
C 0.92688 -3.92970 -3.66574
H -0.01390 -3.40566 -3.57412
C 1.29998 -4.46847 -4.90755
C 2.55152 -5.07918 -5.06154
H 2.86526 -5.45258 -6.02780
C 3.39070 -5.18234 -3.95046
C 4.77291 -5.71911 -3.94046
H 5.38892 -5.24833 -4.73671
H 4.75011 -6.82278 -4.06699
C 5.29103 -5.33226 -2.54575
H 5.88803 -4.39658 -2.63765
H 5.95353 -6.12224 -2.12831
C -5.41634 -2.87129 -1.52709
C -5.06908 -1.81452 -2.52875
C -4.10659 -0.80270 -2.43009
C -3.95392 0.08961 -3.51080
H -3.21188 0.87193 -3.47504
C -4.73636 -0.05568 -4.66734
C -5.64686 -1.11444 -4.77135
H -6.21600 -1.26209 -5.68011
C -5.78068 -1.99301 -3.69513
C -6.59632 -3.22953 -3.67099
H -6.41156 -3.84702 -4.57648
H -7.67349 -2.96929 -3.59007
C -6.10431 -3.95105 -2.40696
H -6.94155 -4.46840 -1.88871
H -5.36239 -4.72429 -2.71055
C -4.30734 -3.46437 -0.70524
C -2.96940 -3.74498 -1.05846
C -2.20957 -4.58891 -0.22519
H -1.20005 -4.85284 -0.51447
C -2.70502 -5.04039 0.94235
C -4.04025 -4.59964 1.40148
H -4.42092 -4.88968 2.37203
C -4.79081 -3.81001 0.53625
C -6.14737 -3.25992 0.78597
H -6.17136 -2.66588 1.72450
H -6.88810 -4.08690 0.82984
C -6.39345 -2.36319 -0.43987
H -6.14549 -1.31230 -0.16897

H -7.45738 -2.39682 -0.76223
C -3.94081 5.34286 -0.68360
C -4.26108 4.09818 0.08406
C -4.10559 2.75434 -0.29002
C -4.72491 1.77341 0.50251
C -5.40774 2.15359 1.66889
C -5.41034 3.49711 2.07858
H -5.88353 3.77996 3.01026
C -4.83160 4.42384 1.29759
C -4.73180 5.87875 1.58280
H -5.74110 6.34225 1.54428
H -4.25754 6.06006 2.57143
C -3.84099 6.40182 0.44117
H -2.79139 6.46505 0.80787
H -4.15525 7.41680 0.11231
C -2.73618 5.32516 -1.57036
C -1.41783 4.69617 -1.33178
C -0.39827 4.95491 -2.26638
H 0.57562 4.50878 -2.13975
C -0.64398 5.69623 -3.36522
C -1.92885 6.20319 -3.62294
H -2.12142 6.75859 -4.53188
C -2.95573 5.97970 -2.69985
C -4.36681 6.41460 -2.84762
H -4.78671 6.08074 -3.82102
H -4.43457 7.51967 -2.75183
C -5.06991 5.71389 -1.67725
H -5.85390 6.36298 -1.22867
H -5.56257 4.79110 -2.05974
H -2.11669 -5.69982 1.56233
H -5.90585 1.40527 2.26775
H 1.43796 4.70482 3.59529
H 5.59637 -2.14045 2.51037
H 0.63586 -4.37679 -5.75709
H -4.60759 0.62874 -5.49572
H 0.14795 5.86632 -4.08343
H 5.82314 1.22998 -4.45873
C 0.07945 -0.25220 2.21739
C 1.18104 0.32758 3.01533
C -1.30575 -0.17995 2.73778
C 2.48731 0.41816 2.46747
C 1.02933 0.74391 4.35592
C -2.19810 -1.26709 2.63805
C -1.80169 1.02900 3.28000
C 3.54977 0.91897 3.19599

H 2.66104 0.06978 1.46102
C 2.08761 1.26142 5.09903
H 0.06181 0.66631 4.83874
C -3.47354 -1.20860 3.17859
H -1.87585 -2.17144 2.13944
C -3.07641 1.10006 3.82692
H -1.17758 1.91489 3.26991
C 3.36264 1.35203 4.51957
H 4.54017 0.98954 2.75741
H 1.91367 1.57755 6.12123
C -3.88736 -0.03481 3.80708
H -4.13961 -2.06093 3.14700
H -3.44570 2.01359 4.27690
H 1.88183 -5.11936 2.08405
C 1.16340 -4.37626 1.71616
Si 0.88865 -3.07783 3.04159
H 0.24266 -4.90177 1.45409
H 1.54947 -3.92876 0.79875
C -0.39599 -3.59113 4.31017
C 2.47835 -2.57831 3.90912
H 0.36423 -1.70107 2.30510
C -0.73721 -2.75192 5.39027
C -1.01840 -4.85124 4.24059
H 3.13680 -1.99279 3.26711
H 2.28595 -1.98781 4.80937
H 3.00527 -3.49378 4.20934
C -1.67551 -3.14495 6.34307
H -0.26854 -1.77591 5.48843
C -1.95563 -5.25192 5.19438
H -0.75441 -5.54151 3.44524
H -1.92918 -2.47739 7.16214
C -2.29244 -4.39490 6.24284
H -2.42111 -6.23095 5.11783
H -3.02826 -4.69977 6.98198
H -4.66604 0.72886 0.23208
H 0.82375 4.33286 1.22723
H 4.86279 -0.90462 0.49233
N -5.19554 0.00460 4.46060
O -5.51096 1.03868 5.05450
O -5.90661 -0.99985 4.38685
O 4.46385 1.82848 5.14904
C 4.33757 2.30928 6.48119
H 3.62052 3.13739 6.54143
H 4.02960 1.51080 7.16791
H 5.32911 2.66679 6.76266

TS-5

Rh -0.08704 -0.00003 0.07542
Rh 0.09650 0.24826 -2.55745
P 2.75820 0.75548 -1.01158
P 0.66365 -2.63782 -1.40650
P -2.73594 -0.48181 -1.53479
P -0.47957 2.89426 -1.11790
O 1.85751 0.74240 0.22659
O 2.12420 0.73113 -2.39269
O 3.71131 2.09086 -1.01523
O 3.78531 -0.51443 -0.79274
O 0.66758 -1.96046 -0.02951
O 0.58140 -1.79358 -2.66146
O 2.01018 -3.55462 -1.62299
O -0.56165 -3.73465 -1.29244
O -1.99728 -0.86294 -0.24357
O -1.93958 -0.20431 -2.79319
O -3.66574 0.79094 -1.05692
O -3.78867 -1.66168 -1.98659
O -0.76554 1.98211 0.08370
O -0.43401 2.28103 -2.50688
O -1.62834 4.07030 -1.20429
O 0.91504 3.67109 -0.73879
C 6.53724 1.01550 -0.74895
C 5.66172 1.67106 0.27321
C 4.39152 2.25335 0.11857
C 3.84756 2.93454 1.15546
C 4.51623 3.02148 2.39168
C 5.73451 2.35347 2.58031
H 6.23870 2.37477 3.53663
C 6.27543 1.65334 1.50403
C 7.50184 0.81880 1.50806
H 7.46657 0.06509 2.32402
H 8.39917 1.46570 1.61200
C 7.46551 0.13833 0.12780
H 8.48736 0.02703 -0.29699
H 7.02583 -0.87829 0.24391
C 5.87000 0.23444 -1.83675
C 4.67512 -0.63536 -1.77918
C 4.36730 -1.38658 -2.92545
H 3.49178 -2.01668 -2.94212
C 5.13348 -1.25378 -4.09480
C 6.19526 -0.34012 -4.16003
H 6.74549 -0.21096 -5.08318

C 6.52986 0.40572 -3.02916
C 7.59218 1.43648 -2.93824
H 7.47403 2.20086 -3.73626
H 8.59164 0.95543 -3.00466
C 7.36759 2.05340 -1.54818
H 8.33333 2.30884 -1.05903
H 6.78484 2.99472 -1.66958
C 0.88932 -6.40193 -1.52753
C 1.65460 -5.62291 -0.49684
C 2.26482 -4.36137 -0.59545
C 3.05861 -3.91055 0.47606
C 3.17321 -4.67908 1.64540
C 2.46865 -5.88460 1.76679
H 2.50218 -6.44935 2.68946
C 1.69923 -6.32537 0.68795
C 0.80041 -7.50624 0.66998
H 1.40262 -8.43988 0.66291
H 0.11003 -7.49323 1.54086
C 0.02346 -7.33650 -0.64749
H -0.95199 -6.84712 -0.42458
H -0.18223 -8.31856 -1.12738
C 0.06985 -5.65592 -2.54193
C -0.68041 -4.47829 -2.38998
C -1.44906 -4.02107 -3.47729
H -2.03409 -3.11648 -3.38586
C -1.41918 -4.70377 -4.70341
C -0.59786 -5.82721 -4.86242
H -0.53390 -6.32937 -5.81911
C 0.15020 -6.27403 -3.77149
C 1.13110 -7.38553 -3.77893
H 1.85284 -7.27440 -4.61672
H 0.59865 -8.35816 -3.85006
C 1.83930 -7.24033 -2.42302
H 2.79460 -6.69004 -2.58147
H 2.08492 -8.23431 -1.98837
C -6.46858 -0.30807 -1.23701
C -5.75946 0.64392 -2.15212
C -4.48815 1.21453 -2.01252
C -4.05557 2.14782 -2.97359
H -3.08023 2.60319 -2.88958
C -4.86962 2.46007 -4.07415
C -6.10513 1.82040 -4.23817
H -6.71720 2.02261 -5.10766
C -6.52075 0.90319 -3.27097
C -7.74624 0.07193 -3.32021

H -7.83332 -0.45204 -4.29648
H -8.63999 0.70630 -3.13757
C -7.53707 -0.93230 -2.17606
H -8.49583 -1.15669 -1.65869
H -7.15277 -1.88362 -2.60944
C -5.66977 -1.37192 -0.53894
C -4.55250 -2.10328 -0.98836
C -4.17362 -3.26968 -0.29474
H -3.34558 -3.86313 -0.66041
C -4.80264 -3.63905 0.83644
C -5.87049 -2.78275 1.40211
H -6.33414 -3.02190 2.35002
C -6.24632 -1.65532 0.67765
C -7.27156 -0.65175 1.06025
H -7.04493 -0.21295 2.05594
H -8.27887 -1.12065 1.06000
C -7.15561 0.40879 -0.04973
H -6.50968 1.23946 0.31452
H -8.14973 0.83149 -0.31450
C -0.06230 6.59660 -0.41584
C -0.99482 5.73893 0.38431
C -1.82519 4.68592 -0.03755
C -2.76669 4.18387 0.87706
C -2.78475 4.66961 2.19592
C -1.84280 5.62717 2.60601
H -1.81896 5.95819 3.63652
C -0.97492 6.12443 1.70962
C 0.10690 7.10533 1.98519
H -0.33286 8.10152 2.20680
H 0.74804 6.76664 2.82761
C 0.90308 7.13401 0.66830
H 1.77817 6.45204 0.76666
H 1.28341 8.15497 0.44475
C 0.65591 5.95490 -1.55978
C 1.24934 4.60224 -1.63264
C 1.96445 4.26824 -2.79860
H 2.40323 3.28793 -2.90506
C 2.05935 5.14303 -3.82009
C 1.42188 6.39408 -3.76311
H 1.47240 7.06836 -4.60846
C 0.71265 6.75472 -2.61260
C -0.05250 8.01154 -2.42220
H -0.75985 8.17831 -3.26317
H 0.64709 8.86955 -2.32547
C -0.80890 7.76379 -1.10994

H -0.85469 8.68585 -0.48958
H -1.85258 7.46132 -1.35420
H -4.49059 -4.53059 1.36467
H -3.49734 4.27677 2.90848
H 4.06916 3.57011 3.21028
H 3.76768 -4.31548 2.47351
H -2.00149 -4.33812 -5.53912
H -4.52607 3.16925 -4.81595
H 2.60036 4.86320 -4.71504
H 4.86781 -1.82594 -4.97422
C -0.27752 -0.01068 2.21397
C 1.00856 0.09462 2.92272
C -1.48800 0.59113 2.87812
C 2.03478 -0.85140 2.67922
C 1.31780 1.16189 3.78962
C -2.62087 0.95183 2.12233
C -1.58564 0.69189 4.28643
C 3.24679 -0.79261 3.33758
H 1.85012 -1.63835 1.95761
C 2.54258 1.24074 4.45093
H 0.60332 1.96385 3.93656
C -3.78981 1.39807 2.73217
H -2.59727 0.87048 1.04981
C -2.73763 1.16046 4.90968
H -0.75816 0.36893 4.90652
C 3.50267 0.24081 4.25724
H 4.01878 -1.53538 3.16277
H 2.73390 2.08133 5.10802
C -3.83082 1.50830 4.11827
H -4.65954 1.66073 2.14295
H -2.80741 1.24223 5.98756
H -3.64328 -3.23535 3.55503
C -3.06431 -2.31625 3.40095
Si -1.26285 -2.75080 3.09298
H -3.21100 -1.67196 4.27076
H -3.46926 -1.78948 2.53256
C -0.29951 -3.00751 4.68414
C -1.05206 -4.17830 1.89989
H -0.62538 -1.44086 2.36999
C 0.92903 -3.69634 4.67244
C -0.77455 -2.54438 5.92736
H -1.40617 -3.91920 0.90096
H 0.00123 -4.44668 1.80456
H -1.59866 -5.05708 2.26571
C 1.65353 -3.91356 5.84340

H 1.32402 -4.07388 3.73512
C -0.05498 -2.75985 7.10357
H -1.72281 -2.01709 5.98435
H 2.59834 -4.44917 5.80457
C 1.16241 -3.44289 7.06304
H -0.44591 -2.39859 8.05120
H 1.72362 -3.61140 7.97839
H -3.44421 3.39410 0.58003
H 2.87729 3.40173 1.04905
H 3.55145 -2.94894 0.41942
N -5.06248 1.98551 4.76639
O -6.01991 2.25740 4.04308
O -5.06074 2.08294 5.99394
O 4.70466 0.19078 4.89025
C 4.94669 1.10463 5.95097
H 4.20191 0.99134 6.74910
H 5.93727 0.85826 6.33678
H 4.94361 2.14576 5.60500

TS-6

Rh 0.05318 -0.03205 0.10444
Rh 0.11685 0.43177 -2.50883
P 2.16605 2.08078 -0.89549
P 1.96465 -1.89070 -1.54920
P -2.06871 -1.54597 -1.57895
P -1.81480 2.28689 -0.90090
O 1.34213 1.58244 0.29317
O 1.62315 1.89538 -2.29933
O 2.44730 3.68505 -0.72008
O 3.63973 1.35971 -0.69441
O 1.67515 -1.35274 -0.13944
O 1.54918 -1.08410 -2.76084
O 3.57668 -2.12953 -1.76361
O 1.30825 -3.40100 -1.50734
O -1.29714 -1.60384 -0.25060
O -1.35208 -1.03695 -2.81763
O -3.42598 -0.69135 -1.25044
O -2.60121 -3.05465 -1.97664
O -1.54570 1.31706 0.25468
O -1.27898 1.96175 -2.28677
O -3.42663 2.48110 -1.15041
O -1.26387 3.73874 -0.35337
C 5.42507 3.94181 -0.31206
C 4.31909 4.02840 0.69330
C 2.93047 3.99176 0.48410

C 2.09812 4.16146 1.54092
C 2.60799 4.31622 2.84448
C 3.99096 4.26358 3.06896
H 4.39207 4.32727 4.07205
C 4.83157 4.09612 1.96870
C 6.30304 3.91066 1.99068
H 6.59818 3.12558 2.71756
H 6.80090 4.87348 2.23459
C 6.62230 3.47032 0.55155
H 7.59628 3.88310 0.20815
H 6.69104 2.35887 0.52869
C 5.20479 3.06435 -1.50563
C 4.53677 1.74645 -1.60267
C 4.66155 1.04618 -2.81296
H 4.17679 0.08919 -2.93756
C 5.33025 1.62220 -3.90558
C 5.85507 2.92079 -3.82652
H 6.32669 3.37005 -4.69101
C 5.76346 3.62737 -2.62643
C 6.24847 5.00712 -2.37687
H 5.82582 5.71734 -3.11979
H 7.35905 5.02793 -2.40662
C 5.72892 5.31550 -0.96252
H 6.46335 5.91808 -0.38411
H 4.79120 5.90960 -1.05192
C 3.74689 -5.19176 -1.81422
C 4.09093 -4.21399 -0.72807
C 4.14129 -2.81048 -0.76884
C 4.69069 -2.12394 0.32952
C 5.10335 -2.82979 1.47083
C 4.93087 -4.21900 1.54014
H 5.19536 -4.76119 2.43865
C 4.41002 -4.88775 0.43075
C 4.11738 -6.33951 0.33062
H 5.06781 -6.91409 0.29862
H 3.48651 -6.68221 1.17812
C 3.35838 -6.45245 -1.00268
H 2.26492 -6.44264 -0.79079
H 3.60144 -7.40143 -1.52941
C 2.69502 -4.81681 -2.81705
C 1.51494 -4.07949 -2.63360
C 0.62093 -3.95530 -3.71374
H -0.30064 -3.40293 -3.59508
C 0.94176 -4.50462 -4.96550
C 2.16853 -5.15451 -5.15462

H 2.44252 -5.53765 -6.12912
C 3.03560 -5.28412 -4.06793
C 4.39950 -5.86508 -4.09723
H 5.00695 -5.41476 -4.91169
H 4.33755 -6.96750 -4.22152
C 4.97023 -5.49464 -2.71879
H 5.59524 -4.57954 -2.82970
H 5.61804 -6.30588 -2.31946
C -5.62368 -2.73195 -1.33339
C -5.28100 -1.67701 -2.33860
C -4.28995 -0.69078 -2.26475
C -4.15461 0.20980 -3.34085
H -3.39412 0.97448 -3.32344
C -4.98048 0.09578 -4.47039
C -5.91960 -0.93957 -4.55259
H -6.52408 -1.06350 -5.44188
C -6.03744 -1.82566 -3.48074
C -6.88461 -3.04027 -3.43852
H -6.74692 -3.65449 -4.35454
H -7.95111 -2.75228 -3.31928
C -6.36984 -3.78543 -2.19749
H -7.20283 -4.28453 -1.65507
H -5.65967 -4.57553 -2.53194
C -4.50525 -3.36135 -0.55246
C -3.18841 -3.67603 -0.95198
C -2.42464 -4.54663 -0.15075
H -1.43373 -4.83695 -0.47672
C -2.89242 -4.99230 1.03024
C -4.19922 -4.51879 1.53686
H -4.55560 -4.80659 2.51722
C -4.95683 -3.70340 0.70193
C -6.29037 -3.12094 0.99853
H -6.27114 -2.53610 1.94282
H -7.05069 -3.92903 1.05715
C -6.55011 -2.20603 -0.21097
H -6.26283 -1.16510 0.06035
H -7.62442 -2.20621 -0.49885
C -3.89785 5.42327 -0.46986
C -4.20605 4.18827 0.31805
C -4.09677 2.84084 -0.05810
C -4.68480 1.87304 0.77370
C -5.29382 2.26974 1.97503
C -5.25243 3.61552 2.37616
H -5.66900 3.91182 3.33027
C -4.70223 4.52873 1.55962

C -4.55861 5.98227 1.83278
H -5.55905 6.46593 1.84493
H -4.02968 6.15745 2.79443
C -3.71834 6.48340 0.64399
H -2.64999 6.52664 0.95524
H -4.02879 7.50358 0.32783
C -2.73940 5.37648 -1.41576
C -1.42312 4.72162 -1.24058
C -0.44476 4.96194 -2.22295
H 0.52586 4.49814 -2.14293
C -0.72920 5.70739 -3.30961
C -2.01528 6.23814 -3.50599
H -2.24074 6.79723 -4.40507
C -3.00046 6.03493 -2.53414
C -4.40755 6.49980 -2.61300
H -4.88214 6.17362 -3.56360
H -4.44676 7.60623 -2.51694
C -5.06666 5.81625 -1.40737
H -5.81361 6.48262 -0.92239
H -5.59683 4.90338 -1.76246
H -2.30160 -5.67229 1.62527
H -5.77029 1.53234 2.60452
H 1.92968 4.43018 3.67997
H 5.51188 -2.29345 2.31747
H 0.25690 -4.39204 -5.79584
H -4.86448 0.78597 -5.29588
H 0.03039 5.86257 -4.06518
H 5.39700 1.07629 -4.83770
C 0.06752 -0.24857 2.22718
C 1.25966 0.26246 2.98960
C -1.27038 -0.11473 2.83717
C 2.53064 0.28106 2.37175
C 1.19415 0.65365 4.34605
C -2.24973 -1.12168 2.67451
C -1.66251 1.04593 3.53492
C 3.66817 0.69820 3.05419
H 2.62730 -0.05598 1.35178
C 2.31953 1.08873 5.03914
H 0.25403 0.60431 4.88042
C -3.48300 -1.04065 3.29393
H -2.01613 -1.98623 2.06578
C -2.91167 1.14910 4.14706
H -0.99506 1.89936 3.57789
C 3.54682 1.10157 4.38192
H 4.63723 0.71087 2.57033

H 2.26417 1.39196 6.07740
C -3.81011 0.07761 4.07773
H -4.20472 -1.84620 3.21265
H -3.16743 2.06215 4.67286
H 1.71873 -5.18285 2.03648
C 1.03080 -4.40086 1.69208
Si 0.81402 -3.12283 3.04775
H 0.08662 -4.87747 1.42004
H 1.43104 -3.94528 0.78495
C -0.49090 -3.60821 4.30679
C 2.43037 -2.72745 3.92426
H 0.33219 -1.72323 2.34653
C -0.83449 -2.75835 5.37697
C -1.11926 -4.86572 4.24375
H 3.13474 -2.19696 3.28212
H 2.27412 -2.12089 4.82097
H 2.89216 -3.67394 4.23492
C -1.77950 -3.13910 6.32803
H -0.36343 -1.78316 5.46864
C -2.06320 -5.25441 5.19580
H -0.85423 -5.56325 3.45483
H -2.03331 -2.46336 7.14035
C -2.40095 -4.38721 6.23562
H -2.53211 -6.23233 5.12534
H -3.14000 -4.68394 6.97510
H -4.65849 0.82618 0.50510
H 1.02533 4.13744 1.39414
H 4.76126 -1.04401 0.31400
O -5.01440 0.03003 4.71086
C -5.34659 1.07676 5.61034
H -5.44363 2.04216 5.09818
H -6.31042 0.80400 6.04359
H -4.59917 1.16978 6.40914
N 4.74653 1.51837 5.12416
O 4.58713 1.98146 6.25363
O 5.83695 1.37152 4.57384

TS-3'

Rh -0.01177 -0.01845 0.04102
Rh 0.25913 0.31092 -2.56924
P 2.57305 1.51971 -0.83669
P 1.50622 -2.33098 -1.42282
P -2.33502 -1.14918 -1.73464
P -1.13623 2.66332 -1.16057
O 1.61718 1.24912 0.33096

O 2.07647 1.30664 -2.25738
O 3.09429 3.07434 -0.80359
O 3.89237 0.59584 -0.50010
O 1.26206 -1.68956 -0.05010
O 1.27683 -1.51560 -2.67925
O 3.05301 -2.86131 -1.56451
O 0.60213 -3.70824 -1.38052
O -1.59538 -1.34906 -0.40225
O -1.55583 -0.67663 -2.94531
O -3.57236 -0.14615 -1.32060
O -3.04416 -2.54801 -2.22887
O -1.22337 1.68551 0.02098
O -0.81750 2.11234 -2.53915
O -2.56920 3.45438 -1.32438
O -0.05845 3.80699 -0.69468
C 6.07406 2.84634 -0.23469
C 4.94949 3.23230 0.67491
C 3.59154 3.44176 0.37697
C 2.78571 3.97902 1.32368
C 3.27788 4.27336 2.60933
C 4.60161 3.95660 2.94499
H 4.97539 4.12661 3.94624
C 5.41712 3.40593 1.95727
C 6.81902 2.94393 2.10936
H 6.91015 2.21768 2.94532
H 7.48857 3.81515 2.27412
C 7.11244 2.26724 0.75775
H 8.16335 2.43928 0.43678
H 6.95822 1.16980 0.86794
C 5.76893 1.91017 -1.36106
C 4.86881 0.73810 -1.39679
C 4.89994 -0.06734 -2.54665
H 4.24444 -0.92037 -2.62851
C 5.70986 0.27915 -3.64024
C 6.47365 1.45520 -3.62625
H 7.05497 1.73629 -4.49498
C 6.47085 2.26250 -2.48791
C 7.18747 3.54986 -2.32148
H 6.94108 4.25174 -3.14717
H 8.28298 3.36918 -2.27782
C 6.66059 4.07615 -0.97671
H 7.46276 4.59006 -0.40262
H 5.85514 4.81819 -1.17947
C 2.70411 -5.90381 -1.59731
C 3.18513 -4.99174 -0.50593

C 3.44794 -3.61190 -0.53872
C 4.04196 -3.01157 0.58732
C 4.29962 -3.76952 1.74104
C 3.93682 -5.12211 1.78976
H 4.08020 -5.69662 2.69583
C 3.35864 -5.70444 0.66059
C 2.80925 -7.07944 0.56203
H 3.63821 -7.81919 0.56428
H 2.10033 -7.28453 1.39311
C 2.07817 -7.06883 -0.79161
H 0.99822 -6.86455 -0.61077
H 2.16425 -8.05044 -1.30748
C 1.76256 -5.36242 -2.63447
C 0.72360 -4.42758 -2.49453
C -0.09321 -4.15445 -3.60838
H -0.89883 -3.43860 -3.52774
C 0.16655 -4.76490 -4.84547
C 1.26222 -5.62569 -4.98811
H 1.49686 -6.06071 -5.95104
C 2.05563 -5.89557 -3.87157
C 3.29662 -6.70667 -3.85432
H 3.99774 -6.37950 -4.65214
H 3.04595 -7.78260 -3.97330
C 3.88198 -6.42551 -2.46159
H 4.65992 -5.63442 -2.55883
H 4.36656 -7.33246 -2.03744
C -6.00521 -1.90375 -1.58833
C -5.51891 -0.81169 -2.49209
C -4.43028 0.05201 -2.31795
C -4.19729 1.04998 -3.28346
H -3.36458 1.72815 -3.17631
C -5.01941 1.14443 -4.41762
C -6.05784 0.22454 -4.61128
H -6.66646 0.26350 -5.50551
C -6.27689 -0.75559 -3.64148
C -7.26117 -1.86013 -3.71898
H -7.18222 -2.39675 -4.68907
H -8.28830 -1.46078 -3.57728
C -6.85703 -2.77482 -2.55246
H -7.75131 -3.22051 -2.06377
H -6.23947 -3.60842 -2.95775
C -4.99333 -2.73709 -0.85477
C -3.71581 -3.17160 -1.26116
C -3.08580 -4.20751 -0.54362
H -2.12644 -4.58072 -0.87816

C -3.64015 -4.71464 0.57335
C -4.90363 -4.14564 1.09591
H -5.32422 -4.48715 2.03248
C -5.52185 -3.14895 0.34678
C -6.77334 -2.42575 0.68679
H -6.69354 -1.94181 1.68419
H -7.63465 -3.12714 0.66074
C -6.88448 -1.37258 -0.43085
H -6.47321 -0.40782 -0.05659
H -7.94264 -1.20697 -0.73067
C -1.85581 6.32268 -0.47698
C -2.54623 5.22732 0.27791
C -3.00346 3.97766 -0.17675
C -3.80547 3.21536 0.69030
C -4.04580 3.67095 1.99799
C -3.45555 4.86294 2.44778
H -3.59244 5.18293 3.47296
C -2.72051 5.59690 1.59655
C -1.99340 6.85131 1.92190
H -2.71889 7.67372 2.10136
H -1.33477 6.71325 2.80657
C -1.16039 7.11660 0.65549
H -0.13160 6.72164 0.81742
H -1.08306 8.20495 0.43986
C -0.90883 5.92236 -1.56274
C 0.05458 4.80063 -1.57656
C 0.90651 4.68940 -2.69219
H 1.61307 3.87581 -2.75718
C 0.80900 5.55672 -3.71976
C -0.16506 6.56966 -3.72177
H -0.25901 7.23104 -4.57351
C -1.01990 6.70634 -2.62297
C -2.12911 7.68423 -2.50127
H -2.79812 7.63976 -3.38771
H -1.71800 8.70881 -2.37417
C -2.86411 7.22199 -1.23558
H -3.21778 8.08830 -0.63430
H -3.75538 6.62770 -1.54028
H -3.13597 -5.49985 1.12146
H -4.65232 3.08397 2.67362
H 2.61958 4.70095 3.35423
H 4.73919 -3.29813 2.61019
H -0.45661 -4.53806 -5.70066
H -4.82804 1.90708 -5.16124
H 1.46224 5.44666 -4.57596

H 5.70485 -0.34290 -4.52588
C -0.27745 -0.12408 2.13636
C 0.94650 0.21825 2.89157
C -1.59682 0.13243 2.77201
C 2.11242 -0.56479 2.76105
C 1.02552 1.39430 3.66842
C -2.77411 0.13237 1.99403
C -1.74200 0.27547 4.17278
C 3.27812 -0.24062 3.43327
H 2.08666 -1.42550 2.10557
C 2.19336 1.72802 4.34569
H 0.17346 2.06194 3.72760
C -4.02744 0.26330 2.58274
H -2.70653 0.01376 0.92598
C -2.99318 0.42893 4.76116
H -0.86782 0.24015 4.81199
C 3.32481 0.90345 4.25028
H 4.16160 -0.86139 3.33039
H 2.23258 2.62586 4.95416
C -4.14264 0.41788 3.96511
H -4.91552 0.25655 1.95952
H -3.07985 0.54776 5.83625
H -2.76724 -4.15350 3.45881
C -2.41071 -3.12580 3.31732
Si -0.55868 -3.13194 2.98773
H -2.68417 -2.54895 4.20461
H -2.94306 -2.68641 2.46951
C 0.43588 -3.18711 4.58267
C -0.03387 -4.46169 1.77211
H -0.26111 -1.73316 2.30293
C 1.72238 -3.75829 4.60934
C -0.06746 -2.66575 5.79141
H -0.35438 -4.21477 0.75836
H 1.05327 -4.56345 1.74219
H -0.45674 -5.43220 2.06033
C 2.47816 -3.80329 5.78041
H 2.13731 -4.18511 3.70265
C 0.68309 -2.70629 6.96705
H -1.06181 -2.22867 5.82150
H 3.46764 -4.25269 5.77040
C 1.95982 -3.27255 6.96324
H 0.27112 -2.30004 7.88712
H 2.54558 -3.30323 7.87796
H -4.19905 2.26005 0.36805
H 1.74300 4.16615 1.10836

H 4.26687 -1.95285 0.58237
C -5.50617 0.49566 4.59976
F -5.99731 -0.74003 4.85337
F -5.47832 1.15949 5.77526
F -6.39806 1.11666 3.79937
C 4.51950 1.22255 4.97291
N 5.49142 1.47819 5.55988

TS-4'

Rh -0.01585 -0.02240 0.06489
Rh -0.10705 0.47340 -2.53354
P 2.16631 1.94855 -1.05833
P 1.64910 -1.97676 -1.73344
P -2.35549 -1.36258 -1.47039
P -1.78048 2.44167 -0.78131
O 1.39560 1.49200 0.18295
O 1.50866 1.82401 -2.41928
O 2.57400 3.52523 -0.88876
O 3.59449 1.12149 -0.97877
O 1.49440 -1.43795 -0.30245
O 1.19769 -1.13080 -2.90560
O 3.22326 -2.30747 -2.06477
O 0.90609 -3.44503 -1.66003
O -1.49048 -1.49168 -0.20558
O -1.69893 -0.87630 -2.75087
O -3.62504 -0.43171 -1.02318
O -3.00682 -2.82676 -1.85044
O -1.50175 1.43426 0.34160
O -1.37665 2.08944 -2.20465
O -3.38481 2.76330 -0.90744
O -1.07061 3.83340 -0.26638
C 5.58671 3.56031 -0.70932
C 4.56918 3.70887 0.37864
C 3.16974 3.77857 0.27660
C 2.43548 3.99792 1.39496
C 3.05375 4.10211 2.65590
C 4.44209 3.94531 2.77271
H 4.92363 3.97676 3.74071
C 5.18103 3.72404 1.61094
C 6.63247 3.43196 1.51869
H 6.92264 2.61538 2.21245
H 7.21693 4.35197 1.73408
C 6.80930 2.99245 0.05489
H 7.78130 3.34106 -0.35818
H 6.79670 1.87943 0.01544

C 5.21445 2.71984 -1.89157
C 4.44363 1.45701 -1.95160
C 4.41700 0.77454 -3.17849
H 3.85146 -0.13972 -3.27814
C 5.03794 1.32072 -4.31371
C 5.66434 2.57455 -4.25953
H 6.09676 3.00528 -5.15346
C 5.72227 3.26199 -3.04646
C 6.32311 4.59935 -2.81999
H 5.89659 5.35020 -3.51949
H 7.42669 4.54190 -2.93620
C 5.93742 4.91981 -1.36625
H 6.75543 5.45945 -0.84021
H 5.04069 5.58018 -1.37579
C 3.20009 -5.36993 -2.20322
C 3.68342 -4.44757 -1.12184
C 3.81231 -3.04863 -1.12915
C 4.48346 -2.42809 -0.05954
C 4.94498 -3.19172 1.02452
C 4.69882 -4.57074 1.06993
H 5.00251 -5.15494 1.92884
C 4.05335 -5.17411 -0.01098
C 3.67210 -6.60382 -0.12678
H 4.58312 -7.22792 -0.24954
H 3.09046 -6.93779 0.75862
C 2.80678 -6.63349 -1.39824
H 1.73534 -6.57261 -1.10003
H 2.95519 -7.57680 -1.96854
C 2.09423 -4.90805 -3.10692
C 0.97751 -4.11093 -2.81052
C 0.00719 -3.90929 -3.81026
H -0.86770 -3.30881 -3.60451
C 0.19348 -4.44167 -5.09608
C 1.36226 -5.15173 -5.40048
H 1.53460 -5.52187 -6.40289
C 2.30568 -5.35815 -4.39216
C 3.62870 -6.00956 -4.54751
H 4.19232 -5.56892 -5.39799
H 3.49628 -7.10314 -4.69286
C 4.32838 -5.70922 -3.21208
H 4.99057 -4.82465 -3.35130
H 4.96164 -6.56404 -2.88733
C -5.94546 -2.33572 -0.96872
C -5.61631 -1.28071 -1.97879
C -4.56385 -0.35717 -1.96557

C -4.45675 0.55748 -3.03282
H -3.65054 1.27350 -3.06221
C -5.37351 0.51945 -4.09529
C -6.37880 -0.45484 -4.12138
H -7.05760 -0.52127 -4.96184
C -6.46737 -1.35675 -3.05996
C -7.38338 -2.51773 -2.97283
H -7.35635 -3.11808 -3.90781
H -8.41713 -2.16777 -2.76435
C -6.82156 -3.32125 -1.79023
H -7.63959 -3.78085 -1.19291
H -6.19076 -4.14554 -2.19419
C -4.81283 -3.05195 -0.28997
C -3.55229 -3.43576 -0.79584
C -2.78730 -4.37352 -0.07547
H -1.84283 -4.71358 -0.48146
C -3.19426 -4.82457 1.12584
C -4.42828 -4.28810 1.74042
H -4.72862 -4.58178 2.73767
C -5.19084 -3.40150 0.98645
C -6.45688 -2.74326 1.39858
H -6.32728 -2.18636 2.35114
H -7.26211 -3.50231 1.49832
C -6.74678 -1.78112 0.23318
H -6.37070 -0.76857 0.50367
H -7.83749 -1.70444 0.02953
C -3.55109 5.73095 -0.16473
C -3.90336 4.51226 0.63009
C -3.94543 3.16710 0.23069
C -4.57325 2.24499 1.08467
C -5.06376 2.67516 2.32846
C -4.86528 4.00177 2.74499
H -5.18969 4.31650 3.72870
C -4.28393 4.87438 1.90608
C -3.99232 6.30455 2.18540
H -4.94230 6.87404 2.27603
H -3.38259 6.41493 3.10804
C -3.19840 6.75154 0.94437
H -2.11076 6.69549 1.17686
H -3.43955 7.80026 0.66306
C -2.47689 5.60461 -1.19768
C -1.21245 4.83849 -1.13157
C -0.29658 5.00927 -2.18638
H 0.63322 4.46304 -2.18824
C -0.59869 5.79395 -3.24017

C -1.84501 6.43637 -3.32947
H -2.09029 7.02746 -4.20253
C -2.76583 6.30161 -2.28543
C -4.12976 6.88488 -2.24956
H -4.70186 6.61681 -3.16394
H -4.06678 7.98878 -2.13828
C -4.75045 6.23792 -1.00419
H -5.39855 6.95610 -0.45520
H -5.38237 5.37957 -1.32736
H -2.60489 -5.55570 1.65830
H -5.56820 1.97632 2.97914
H 2.45210 4.26035 3.54140
H 5.44965 -2.70601 1.84964
H -0.54869 -4.26881 -5.86450
H -5.27944 1.21991 -4.91488
H 0.11045 5.89559 -4.05181
H 4.98724 0.79066 -5.25594
C 0.11036 -0.30340 2.15772
C 1.33734 0.11872 2.89022
C -1.20063 -0.15770 2.83075
C 2.58887 0.11160 2.23602
C 1.32326 0.46477 4.26186
C -2.17497 -1.17376 2.78843
C -1.54187 1.05235 3.47679
C 3.76004 0.45028 2.90977
H 2.64139 -0.18613 1.19996
C 2.48901 0.80816 4.93530
H 0.39130 0.45488 4.81347
C -3.37298 -1.05246 3.47507
H -1.97502 -2.07334 2.22249
C -2.74491 1.18565 4.15857
H -0.85839 1.89296 3.43802
C 3.71461 0.79472 4.26133
H 4.70744 0.43857 2.38157
H 2.45227 1.07162 5.98789
C -3.65596 0.11665 4.20001
H -4.08455 -1.87032 3.47933
H -2.97835 2.11278 4.67263
H 1.46023 -5.38126 1.80198
C 0.80534 -4.55096 1.51045
Si 0.77471 -3.27409 2.88438
H -0.18524 -4.96138 1.30187
H 1.17147 -4.11581 0.57894
C -0.43465 -3.70757 4.25499
C 2.48002 -2.96118 3.61189

H 0.31167 -1.86218 2.24791
C -0.60186 -2.86897 5.37560
C -1.17632 -4.90282 4.21852
H 3.13426 -2.42906 2.92028
H 2.43226 -2.37961 4.53711
H 2.93907 -3.93051 3.84613
C -1.48800 -3.19754 6.40018
H -0.03518 -1.94427 5.45107
C -2.06250 -5.23943 5.24344
H -1.04685 -5.59445 3.39144
H -1.60585 -2.53124 7.25043
C -2.22636 -4.38203 6.33227
H -2.62225 -6.16950 5.19128
H -2.92087 -4.63696 7.12815
H -4.66607 1.20590 0.79988
H 1.35590 4.05594 1.33064
H 4.61433 -1.35366 -0.05103
C 4.97047 1.07741 5.04249
F 4.84807 2.19527 5.79213
F 5.24545 0.06465 5.89793
F 6.04757 1.23209 4.24769
C -4.85540 0.20776 4.97718
N -5.82730 0.27592 5.61392

TS-5'

Rh -0.20388 0.06083 0.09410
Rh -0.29593 0.50186 -2.51130
P 2.59458 0.36663 -1.29895
P -0.20981 -2.51843 -1.66653
P -3.07766 0.25665 -1.16141
P -0.15503 3.07080 -0.83851
O 1.85415 0.37925 0.04206
O 1.79765 0.54131 -2.58067
O 3.72182 1.55397 -1.34606
O 3.43372 -1.05104 -1.28751
O 0.10212 -1.98432 -0.26093
O -0.24577 -1.56362 -2.84312
O 0.86891 -3.67044 -2.11703
O -1.62565 -3.33673 -1.46408
O -2.27758 -0.36438 -0.00637
O -2.39635 0.48333 -2.49529
O -3.67233 1.63543 -0.48585
O -4.38746 -0.66109 -1.54097
O -0.45615 2.12475 0.33517
O -0.41457 2.58068 -2.25283

O -1.04542 4.44520 -0.68150
O 1.40295 3.53111 -0.62350
C 6.36242 0.11609 -1.62129
C 5.76385 0.78172 -0.42288
C 4.57287 1.51837 -0.32245
C 4.28641 2.15610 0.83769
C 5.14264 2.04763 1.95033
C 6.27955 1.22951 1.88589
H 6.92004 1.10060 2.74822
C 6.55636 0.58064 0.68331
C 7.65699 -0.37748 0.41637
H 7.67146 -1.19148 1.17195
H 8.62974 0.15923 0.40358
C 7.30979 -0.92832 -0.97836
H 8.22628 -1.11359 -1.58061
H 6.77562 -1.89763 -0.85437
C 5.42340 -0.49647 -2.61269
C 4.16118 -1.23718 -2.39068
C 3.60859 -1.91210 -3.49063
H 2.68277 -2.45580 -3.38221
C 4.19145 -1.80171 -4.76384
C 5.31560 -0.98936 -4.97368
H 5.72595 -0.87512 -5.96873
C 5.90095 -0.33369 -3.88980
C 7.08102 0.56441 -3.93506
H 6.92619 1.38774 -4.66520
H 7.99029 -0.01962 -4.19362
C 7.16554 1.10936 -2.49943
H 8.22219 1.22439 -2.17174
H 6.68547 2.11397 -2.47315
C -0.83695 -6.21436 -2.07560
C 0.19137 -5.71629 -1.10132
C 1.05120 -4.60862 -1.19054
C 2.04070 -4.44178 -0.20347
C 2.11332 -5.32773 0.88361
C 1.17922 -6.36466 1.00987
H 1.19265 -7.01308 1.87640
C 0.21411 -6.52534 0.01388
C -0.91688 -7.48630 0.02934
H -0.53774 -8.51930 -0.12415
H -1.48504 -7.41646 0.98201
C -1.78145 -7.02992 -1.15880
H -2.59460 -6.37245 -0.77508
H -2.24998 -7.89562 -1.67661
C -1.58285 -5.21848 -2.91691

C -2.03283 -3.93017 -2.58427
C -2.80125 -3.21903 -3.52566
H -3.15711 -2.22474 -3.29565
C -3.06357 -3.77089 -4.78939
C -2.53332 -5.02051 -5.13436
H -2.69133 -5.42960 -6.12389
C -1.78165 -5.71747 -4.18656
C -1.07521 -7.00282 -4.40248
H -0.44781 -6.96365 -5.31898
H -1.81212 -7.83167 -4.47011
C -0.20142 -7.14033 -3.14602
H 0.82616 -6.78720 -3.39073
H -0.13015 -8.20112 -2.81904
C -6.64074 1.11432 -0.37550
C -5.86281 1.98911 -1.31118
C -4.49564 2.29188 -1.29861
C -3.99611 3.20632 -2.24527
H -2.94645 3.45807 -2.25754
C -4.84974 3.76388 -3.21053
C -6.19860 3.38851 -3.25640
H -6.85165 3.77948 -4.02591
C -6.67775 2.48595 -2.30502
C -8.04220 1.91101 -2.25092
H -8.34120 1.50087 -3.23964
H -8.76495 2.68554 -1.91578
C -7.91051 0.78900 -1.20929
H -8.83072 0.70651 -0.58990
H -7.77427 -0.17633 -1.74789
C -5.99802 -0.13842 0.14871
C -5.10812 -1.02979 -0.48278
C -4.90421 -2.30563 0.07909
H -4.26257 -3.01542 -0.42674
C -5.46806 -2.64464 1.25350
C -6.27019 -1.65064 2.00295
H -6.66833 -1.87904 2.98280
C -6.48502 -0.41179 1.40620
C -7.24078 0.73207 1.97630
H -6.82519 1.03002 2.96308
H -8.31582 0.46700 2.06908
C -7.03414 1.84372 0.93156
H -6.19669 2.49756 1.26552
H -7.94468 2.47272 0.82039
C 1.09833 6.54545 0.01127
C 0.12628 5.81627 0.88882
C -0.95172 4.98512 0.53480

C -1.84517 4.59967 1.54883
C -1.58791 4.96653 2.88101
C -0.42334 5.68270 3.20141
H -0.19555 5.91406 4.23424
C 0.40126 6.07513 2.21669
C 1.68766 6.79966 2.38572
H 1.49456 7.83906 2.72813
H 2.35264 6.27081 3.10237
C 2.28951 6.78799 0.96938
H 3.01130 5.94281 0.89646
H 2.83754 7.73105 0.75209
C 1.51118 5.88168 -1.26318
C 1.79690 4.45138 -1.50429
C 2.25154 4.08659 -2.78642
H 2.45357 3.05185 -3.01802
C 2.38216 5.01233 -3.75778
C 2.03502 6.35469 -3.52942
H 2.10446 7.07768 -4.33205
C 1.58428 6.74446 -2.26388
C 1.12518 8.10368 -1.88597
H 0.35954 8.47808 -2.59939
H 1.99195 8.79790 -1.84455
C 0.51888 7.89173 -0.49206
H 0.74670 8.74668 0.18194
H -0.58783 7.81968 -0.59423
H -5.29094 -3.62398 1.67857
H -2.26535 4.66286 3.66737
H 4.90186 2.56588 2.86918
H 2.86324 -5.18206 1.64995
H -3.64281 -3.21191 -5.51269
H -4.45688 4.45714 -3.94280
H 2.71975 4.71121 -4.74130
H 3.73574 -2.30860 -5.60449
C -0.11342 -0.11471 2.19977
C 1.23965 -0.42430 2.70922
C -1.03563 0.64250 3.09186
C 1.89042 -1.61876 2.33552
C 1.96247 0.50295 3.48728
C -2.17925 1.28094 2.56698
C -0.86762 0.65687 4.49797
C 3.16543 -1.91315 2.78883
H 1.37499 -2.30258 1.67351
C 3.24782 0.20967 3.94062
H 1.52619 1.46638 3.72590
C -3.10186 1.90780 3.39600

H -2.35323 1.27338 1.50430
C -1.77150 1.30138 5.33310
H -0.03197 0.13224 4.94535
C 3.84384 -1.00680 3.61331
H 3.64487 -2.84274 2.49952
H 3.78230 0.92678 4.55398
C -2.90096 1.93410 4.78477
H -3.97896 2.38185 2.96735
H -1.61993 1.30643 6.40784
H -3.97485 -2.64340 3.87136
C -3.21667 -1.88362 3.64506
Si -1.63657 -2.71205 3.05097
H -3.07132 -1.27454 4.54082
H -3.60477 -1.22483 2.86334
C -0.58642 -3.33661 4.47952
C -1.95049 -4.04424 1.76758
H -0.81160 -1.55498 2.34380
C 0.40296 -4.31687 4.26936
C -0.75729 -2.85767 5.79364
H -2.20031 -3.61176 0.79687
H -1.06477 -4.66816 1.62838
H -2.76992 -4.69599 2.09603
C 1.18798 -4.79766 5.31639
H 0.56025 -4.71672 3.27266
C 0.02411 -3.33541 6.84684
H -1.51619 -2.10906 6.00418
H 1.94340 -5.55497 5.12455
C 1.00034 -4.30514 6.60961
H -0.13144 -2.95438 7.85280
H 1.61010 -4.67698 7.42859
H -2.70297 3.98394 1.31158
H 3.38131 2.74197 0.92733
H 2.72358 -3.60342 -0.25639
C 5.18654 -1.39114 4.16767
F 5.82012 -0.35314 4.75398
F 6.00300 -1.87817 3.20562
F 5.07206 -2.36508 5.10266
C -3.84739 2.58980 5.63874
N -4.61705 3.12202 6.32993

TS-6'

Rh 0.18256 0.01034 0.05651
Rh 0.57603 0.49465 -2.51560
P 2.39582 2.12882 -0.62155
P 2.29658 -1.82976 -1.35765

P -1.70475 -1.49286 -1.89134
P -1.54404 2.33282 -1.14695
O 1.39570 1.64824 0.43317
O 2.05874 1.93970 -2.08713
O 2.67986 3.72897 -0.41207
O 3.80753 1.38257 -0.20047
O 1.83351 -1.29256 0.00541
O 2.02768 -1.01977 -2.60836
O 3.92295 -2.06279 -1.36735
O 1.64793 -3.34274 -1.40292
O -1.08571 -1.56781 -0.48513
O -0.85726 -0.95535 -3.03191
O -3.09745 -0.65362 -1.70309
O -2.17625 -2.99921 -2.36290
O -1.44331 1.33070 0.01027
O -0.84451 2.01837 -2.46068
O -3.10790 2.57080 -1.58570
O -1.03242 3.75433 -0.49789
C 5.54368 3.89292 0.56681
C 4.26311 4.01571 1.33409
C 2.94026 4.03966 0.85850
C 1.92865 4.26172 1.73338
C 2.18231 4.40704 3.11062
C 3.48998 4.28622 3.60044
H 3.69041 4.34038 4.66245
C 4.51891 4.06370 2.68559
C 5.94461 3.79417 2.99524
H 6.04347 2.98609 3.75080
H 6.43615 4.72436 3.35222
C 6.51929 3.34736 1.63918
H 7.56245 3.70770 1.50179
H 6.53052 2.23404 1.60832
C 5.53124 3.05518 -0.67445
C 4.84367 1.77304 -0.94241
C 5.14465 1.11364 -2.14466
H 4.65038 0.18599 -2.39276
C 6.01963 1.69435 -3.07704
C 6.58116 2.95915 -2.84755
H 7.21778 3.41480 -3.59501
C 6.30754 3.62657 -1.65258
C 6.79345 4.97348 -1.26633
H 6.54844 5.72381 -2.04866
H 7.88962 4.94205 -1.08714
C 6.02984 5.26659 0.03557
H 6.66837 5.81390 0.76347

H 5.15533 5.91243 -0.20631	C -2.86314 -3.63819 -1.41352
C 4.11888 -5.11974 -1.43762	C -2.17767 -4.49778 -0.53295
C 4.31781 -4.15791 -0.30216	H -1.15129 -4.76879 -0.74639
C 4.35720 -2.75357 -0.31526	C -2.76520 -4.95418 0.58897
C 4.74362 -2.07961 0.85805	C -4.12803 -4.50489 0.94820
C 5.00312 -2.79912 2.03527	H -4.58454 -4.79944 1.88409
C 4.84445 -4.19130 2.05884	C -4.80202 -3.69881 0.03616
H 4.99087 -4.74526 2.97695	C -6.16706 -3.13407 0.18798
C 4.49007 -4.84754 0.87850	H -6.25954 -2.55701 1.13265
C 4.23096 -6.30021 0.71940	H -6.91926 -3.95119 0.15667
H 5.18516 -6.86381 0.79894	C -6.30302 -2.21179 -1.03587
H 3.50249 -6.66458 1.47452	H -6.06132 -1.17007 -0.72604
C 3.64807 -6.39829 -0.70054	H -7.33887 -2.22210 -1.44069
H 2.53661 -6.40442 -0.62792	C -3.60676 5.49182 -0.80447
H 3.96775 -7.33561 -1.20678	C -4.02213 4.22251 -0.12729
C 3.19046 -4.73716 -2.55333	C -3.90353 2.89504 -0.56826
C 1.98998 -4.01106 -2.50234	C -4.60554 1.90146 0.13459
C 1.22968 -3.87968 -3.67975	C -5.33244 2.24771 1.28436
H 0.29630 -3.33457 -3.66526	C -5.30011 3.56567 1.76918
C 1.70214 -4.40984 -4.89105	H -5.80860 3.81783 2.69107
C 2.94922 -5.04655 -4.94106	C -4.64224 4.50676 1.07231
H 3.34110 -5.41362 -5.88085	C -4.49264 5.93720 1.44618
C 3.68223 -5.18377 -3.76070	H -5.47658 6.44991 1.38325
C 5.04645 -5.75146 -3.63488	H -4.06427 6.04023 2.46653
H 5.74158 -5.28356 -4.36493	C -3.51910 6.47991 0.38372
H 5.01199 -6.85247 -3.78043	H -2.48870 6.47139 0.80607
C 5.44485 -5.39589 -2.19327	H -3.76888 7.52568 0.09909
H 6.06721 -4.47258 -2.21616	C -2.35978 5.46670 -1.63095
H 6.05014 -6.20621 -1.73043	C -1.08326 4.76816 -1.36198
C -5.25140 -2.71581 -2.05342	C -0.00895 5.02013 -2.23503
C -4.80978 -1.64467 -3.00162	H 0.93715 4.52351 -2.08476
C -3.84576 -0.64807 -2.80543	C -0.16730 5.82364 -3.30588
C -3.60476 0.27054 -3.84697	C -1.41346 6.40459 -3.59602
H -2.86079 1.04372 -3.73411	H -1.53602 7.01225 -4.48341
C -4.29893 0.16348 -5.06258	C -2.49339 6.18595 -2.73422
C -5.20916 -0.88175 -5.26245	C -3.87407 6.69472 -2.92563
H -5.71024 -0.99916 -6.21466	H -4.25940 6.43154 -3.93437
C -5.43318 -1.78513 -4.22234	H -3.89627 7.79547 -2.77460
C -6.26600 -3.00861 -4.28955	C -4.66611 5.96997 -1.82913
H -6.01997 -3.60934 -5.19172	H -5.44008 6.63310 -1.38367
H -7.34232 -2.73275 -4.28684	H -5.18241 5.09423 -2.28416
C -5.88484 -3.76503 -3.00766	H -2.23111 -5.62303 1.24686
H -6.76755 -4.27901 -2.56740	H -5.89106 1.49014 1.81380
H -5.13315 -4.54390 -3.27000	H 1.36105 4.56442 3.79754
C -4.22024 -3.34409 -1.15986	H 5.28366 -2.27216 2.93803

H 1.11993 -4.29130 -5.79556
H -4.10175 0.86737 -5.86062
H 0.66535 5.98971 -3.97748
H 6.22553 1.18044 -4.00700
C -0.05874 -0.23615 2.14700
C 1.01773 0.24906 3.06444
C -1.47024 -0.11517 2.58154
C 2.35598 0.29076 2.61266
C 0.77795 0.61253 4.40925
C -2.40243 -1.13763 2.31819
C -1.94591 1.05597 3.21135
C 3.39060 0.70908 3.43942
H 2.58339 -0.02633 1.60747
C 1.80402 1.03785 5.24515
H -0.22370 0.55357 4.81627
C -3.69740 -1.07324 2.81243
H -2.08966 -1.99914 1.74286
C -3.24758 1.12601 3.70041
H -1.29282 1.91492 3.31627
C 3.12310 1.08917 4.76476
H 4.40676 0.74206 3.05983
H 1.59452 1.31474 6.27334
C -4.11164 0.03880 3.55071
H -4.37958 -1.90039 2.65130
H -3.58610 2.02192 4.21044
H 1.66717 -5.17124 2.17432
C 1.02266 -4.40020 1.73479
Si 0.55956 -3.15043 3.05442

H 0.14927 -4.89486 1.30522
H 1.55145 -3.91838 0.91094
C -0.92066 -3.68799 4.07844
C 2.01497 -2.73495 4.17141
H 0.16970 -1.75951 2.30754
C -1.53073 -2.82923 5.01421
C -1.43896 -4.99099 3.95351
H 2.80753 -2.19968 3.64534
H 1.71538 -2.13107 5.03259
H 2.43046 -3.67827 4.54992
C -2.63041 -3.23991 5.76638
H -1.15642 -1.81773 5.14822
C -2.53014 -5.41352 4.71481
H -0.97933 -5.69387 3.26389
H -3.10533 -2.54679 6.45509
C -3.13510 -4.53384 5.61382
H -2.90986 -6.42540 4.60086
H -3.99503 -4.85447 6.19573
H -4.57463 0.87113 -0.18988
H 0.90500 4.28479 1.38211
H 4.80466 -0.99931 0.86784
C -5.44106 0.01266 4.24831
F -5.90734 1.25119 4.51882
F -5.35383 -0.63906 5.43738
F -6.38346 -0.63196 3.52545
C 4.18766 1.50449 5.63010
N 5.05224 1.84024 6.33243

References

- [1] Kyriacos C Nicolaou and Tamsyn Montagnon. *Molecules that changed the world*, volume 11. Wiley-VCH Weinheim, 2008.
- [2] Eric N Jacobsen, Andreas Pfaltz, and Hisashi Yamamoto. *Comprehensive asymmetric catalysis: Supplement 1*, volume 1. Springer Science & Business Media, 2003.
- [3] Giovanni Desimoni, Giuseppe Faita, and Karl Anker Jørgensen. Update 1 of: C 2-symmetric chiral bis (oxazoline) ligands in asymmetric catalysis. *Chemical Reviews*, 111(11):PR284–PR437, 2011.
- [4] Dixit Parmar, Erli Sugiono, Sadiya Raja, and Magnus Rueping. Complete field guide to asymmetric binol-phosphate derived brønsted acid and metal catalysis: history and classification by mode of activation; brønsted acidity, hydrogen bonding, ion pairing, and metal phosphates. *Chemical reviews*, 114(18):9047–9153, 2014.
- [5] Subrata Shaw and James D White. Asymmetric catalysis using chiral salen–metal complexes: recent advances. *Chemical reviews*, 119(16):9381–9426, 2019.
- [6] PJ Walsh and MC Kozlowski. *Fundamentals of asymmetric catalysis*, 688 seiten, 2008.
- [7] Taichiro Touge, Hideki Nara, Mitsuhiko Fujiwhara, Yoshihito Kayaki, and Takao Ikariya. Efficient access to chiral benzhydrols via asymmetric transfer hydrogenation of unsymmetrical benzophenones with bifunctional oxo-tethered ruthenium catalysts. *Journal of the American Chemical Society*, 138(32):10084–10087, 2016.
- [8] Qiongzong Jiang, Yutong Jiang, Dengming Xiao, Ping Cao, and Xumu Zhang. Highly enantioselective hydrogenation of simple ketones catalyzed by a rh–penphos complex. *Angewandte Chemie International Edition*, 37(8):1100–1103, 1998.
- [9] Marcel Garbe, Kathrin Junge, Svenja Walker, Zhihong Wei, Haijun Jiao, Anke Spannenberg, Stephan Bachmann, Michelangelo Scalone, and Matthias Beller. Manganese(i)-catalyzed enantioselective hydrogenation of ketones using a defined chiral pnp pincer ligand. *Angewandte Chemie International Edition*, 56(37):11237–11241, 2017.
- [10] Javier Mazuela, J Johan Verendel, Mercedes Coll, Benjamin Schöffner, Armin Börner, Pher G Andersson, Oscar Pamies, and Montserrat Dieguez. Iridium phosphite- oxazoline catalysts for the highly enantioselective hydrogenation of terminal alkenes. *Journal of the American Chemical Society*, 131(34):12344–12353, 2009.
- [11] Javier Mazuela, Per-Ola Norrby, Pher G Andersson, Oscar Pamies, and Montserrat Dieguez. Pyranoside phosphite–oxazoline ligands for the highly versatile and enantioselective iridium-catalyzed hydrogenation of minimally functionalized olefins. a combined theoretical and experimental study. *Journal of the American Chemical Society*, 133(34):13634–13645, 2011.
- [12] Tatiana Besset, Rafael Gramage-Doria, and Joost NH Reek. Remotely controlled iridium-catalyzed asymmetric hydrogenation of terminal 1, 1-diaryl alkenes. *Angewandte Chemie International Edition*, 52(34):8795–8797, 2013.

- [13] Johannes M Wahl, Michael L Conner, and M Kevin Brown. Allenates in enantioselective [2+ 2] cycloadditions: From a mechanistic curiosity to a stereospecific transformation. *Journal of the American Chemical Society*, 140(46):15943–15949, 2018.
- [14] Justin T Mohr, Allen Y Hong, and Brian M Stoltz. Enantioselective protonation. *Nature chemistry*, 1(5):359–369, 2009.
- [15] Quirin M Kainz, Carson D Matier, Agnieszka Bartoszewicz, Susan L Zultanski, Jonas C Peters, and Gregory C Fu. Asymmetric copper-catalyzed cross-couplings induced by visible light. *Science*, 351(6274):681–684, 2016.
- [16] Wen Zhang, Fei Wang, Scott D McCann, Dinghai Wang, Pinhong Chen, Shannon S Stahl, and Guosheng Liu. Enantioselective cyanation of benzylic C–H bonds via copper-catalyzed radical relay. *Science*, 353(6303):1014–1018, 2016.
- [17] Alison E Wendlandt, Prithvi Vangal, and Eric N Jacobsen. Quaternary stereocentres via an enantioconvergent catalytic S_N1 reaction. *Nature*, 556(7702):447–451, 2018.
- [18] Jean-Yves Legros and Jean-Claude Fiaud. Electronic control of enantioselectivity in the palladium-catalyzed asymmetric allylic substitution of trans 4-*t*-butyl-1-vinylcyclohexyl benzoates. *Tetrahedron*, 50(2):465–474, 1994.
- [19] EJ Corey and Christopher J Helal. Novel electronic effects of remote substituents on the oxazaborolidine-catalyzed enantioselective reduction of ketones. *Tetrahedron letters*, 36(50):9153–9156, 1995.
- [20] Huichang Zhang, Feng Xue, TCW Mak, and Kin Shing Chan. Enantioselectivity increases with reactivity: electronically controlled asymmetric addition of diethylzinc to aromatic aldehydes catalyzed by a chiral pyridylphenol. *The Journal of organic chemistry*, 61(23):8002–8003, 1996.
- [21] Hei Lam Wong, Yuan Tian, and Kin Shing Chan. Electronically controlled asymmetric cyclopropanation catalyzed by a new type of chiral 2, 2'-bipyridine. *Tetrahedron Letters*, 41(40):7723–7726, 2000.
- [22] Christophe Werle, Richard Goddard, Petra Philipps, Christophe Farès, and Alois Furstner. Structures of reactive donor/acceptor and donor/donor rhodium carbenes in the solid state and their implications for catalysis. *Journal of the American Chemical Society*, 138(11):3797–3805, 2016.
- [23] Shou-Fei Zhu and Qi-Lin Zhou. Transition-metal-catalyzed enantioselective heteroatom–hydrogen bond insertion reactions. *Accounts of Chemical Research*, 45(8):1365–1377, 2012.
- [24] Hoda Keipour, Virginie Carreras, and Thierry Ollevier. Recent progress in the catalytic carbene insertion reactions into the silicon–hydrogen bond. *Organic & Biomolecular Chemistry*, 15(26):5441–5456, 2017.

- [25] Cristian Soldi, Kellan N Lamb, Richard A Squitieri, Marcos Gonzalez-Lopez, Michael J Di Maso, and Jared T Shaw. Enantioselective intramolecular c–h insertion reactions of donor–donor metal carbenoids. *Journal of the American Chemical Society*, 136(43):15142–15145, 2014.
- [26] Huw ML Davies, Tore Hansen, James Rutberg, and Paul R Bruzinski. Rhodium (ii)(s)-n-(arylsulfonyl) proline catalyzed asymmetric insertions of vinyl- and phenylcarbenoids into the si–h bond. *Tetrahedron letters*, 38(10):1741–1744, 1997.
- [27] Yong-Zhen Zhang, Shou-Fei Zhu, Li-Xin Wang, and Qi-Lin Zhou. Copper-catalyzed highly enantioselective carbenoid insertion into si–h bonds. *Angewandte Chemie International Edition*, 47(44):8496–8498, 2008.
- [28] Yoichi Yasutomi, Hidehiro Suematsu, and Tsutomu Katsuki. Iridium (iii)-catalyzed enantioselective si–h bond insertion and formation of an enantioenriched silicon center. *Journal of the American Chemical Society*, 132(13):4510–4511, 2010.
- [29] Diao Chen, Dong-Xing Zhu, and Ming-Hua Xu. Rhodium (i)-catalyzed highly enantioselective insertion of carbenoid into si–h: Efficient access to functional chiral silanes. *Journal of the American Chemical Society*, 138(5):1498–1501, 2016.
- [30] Stephen Hyde, Janis Veliks, Benoît Liégault, David Grassi, Marc Taillefer, and Véronique Gouverneur. Copper-catalyzed insertion into heteroatom–hydrogen bonds with trifluorodiazalkanes. *Angewandte Chemie International Edition*, 55(11):3785–3789, 2016.
- [31] Michael P Doyle, M Anthony McKervy, Tao Ye, et al. *Modern catalytic methods for organic synthesis with diazo compounds*. Wiley, 1998.
- [32] Alan Ford, Hugues Miel, Aoife Ring, Catherine N Slattery, Anita R Maguire, and M Anthony McKervy. Modern organic synthesis with α -diazocarbonyl compounds. *Chemical reviews*, 115(18):9981–10080, 2015.
- [33] Huw ML Davies and Kuangbiao Liao. Dirhodium tetracarboxylates as catalysts for selective intermolecular c–h functionalization. *Nature Reviews Chemistry*, 3(6):347–360, 2019.
- [34] Bin Xu, Shou-Fei Zhu, Zhi-Chao Zhang, Zhi-Xiang Yu, Yi Ma, and Qi-Lin Zhou. Highly enantioselective s–h bond insertion cooperatively catalyzed by dirhodium complexes and chiral spiro phosphoric acids. *Chemical Science*, 5(4):1442–1448, 2014.
- [35] Corwin Hansch, A Leo, and RW Taft. A survey of hammett substituent constants and resonance and field parameters. *Chemical reviews*, 91(2):165–195, 1991.
- [36] Les A. Dakin, Patricia C. Ong, James S. Panek, Richard J. Staples, and Pericles Stavropoulos. Speciation and mechanistic studies of chiral copper(i) schiff base precursors mediating asymmetric carbenoid insertion reactions of diazoacetates into the si–h bond of silanes. *Organometallics*, 19(15):2896–2908, 2000.

- [37] Maizie Lee, Zhi Ren, Djamaladdin G. Musaev, and Huw M. L. Davies. Rhodium-stabilized diarylcarbenes behaving as donor/acceptor carbenes. *ACS Catalysis*, 10(11):6240–6247, 2020.
- [38] Maizie Lee, Zhi Ren, Djamaladdin G Musaev, and Huw ML Davies. Rhodium-stabilized diarylcarbenes behaving as donor/acceptor carbenes. *ACS Catalysis*, 10(11):6240–6247, 2020.
- [39] Marvin Charton. Steric effects. ii. base-catalyzed ester hydrolysis. *Journal of the American Chemical Society*, 97(13):3691–3693, 1975.
- [40] Marvin Charton. Steric effects. 7. additional v constants. *The Journal of Organic Chemistry*, 41(12):2217–2220, 1976.
- [41] Matthew S Sigman and Jeremie J Miller. Examination of the role of taft-type steric parameters in asymmetric catalysis. *The Journal of organic chemistry*, 74(20):7633–7643, 2009.
- [42] J. Wei, B. Wang, J. Qiu, X. Zhang, H. Chen, and M. Zhang. Preparation of polymerizable benzophenone photoinitiator containing amine group as promoter, CN Patent 102212151, 2011.
- [43] Maria L Carmellino, Giuseppe Pagani, Massimo Pregnotato, Marco Terreni, Vincenzo Caprioli, and Franca Zani. Studies on the insecticidal activities of some new n-benzoyl-n'-arylureas. *Pesticide science*, 45(3):227–236, 1995.
- [44] Siphamandla Sithebe and Patience Molefe. Base free suzuki acylation reactions of sodium (aryl trihydroxyborate) salts: A novel synthesis of substituted aryl ketones. *Journal of Organometallic Chemistry*, 846:305–311, 2017.
- [45] Souya Dohi, Katsuhiko Moriyama, and Hideo Togo. Practical one-pot preparation of ketones from aryl and alkyl bromides with aldehydes and dihalides via grignard reagents. *Tetrahedron*, 68(32):6557–6564, 2012.
- [46] Qing Zhou, Shaohua Wei, and Wei Han. In situ generation of palladium nanoparticles: ligand-free palladium catalyzed pivalic acid assisted carbonylative suzuki reactions at ambient conditions. *The Journal of Organic Chemistry*, 79(3):1454–1460, 2014.
- [47] Lanny S Liebeskind and Jiri Srogl. Thiol ester- boronic acid coupling. a mechanistically unprecedented and general ketone synthesis. *Journal of the American Chemical Society*, 122(45):11260–11261, 2000.
- [48] Mingyou Hu, Zhengbiao He, Bing Gao, Lingchun Li, Chuanfa Ni, and Jinbo Hu. Copper-catalyzed gem-difluoroolefination of diazo compounds with tmscf₃ via c–f bond cleavage. *Journal of the American Chemical Society*, 135(46):17302–17305, 2013.
- [49] Myunghee Lee, Sangwon Ko, and Sukbok Chang. Highly selective and practical hydrolytic oxidation of organosilanes to silanols catalyzed by a ruthenium complex. *Journal of the American Chemical Society*, 122(48):12011–12012, 2000.

- [50] Nirmalya Moitra, Shun Ichii, Toshiyuki Kamei, Kazuyoshi Kanamori, Yang Zhu, Kazuyuki Takeda, Kazuki Nakanishi, and Toyoshi Shimada. Surface functionalization of silica by si–h activation of hydrosilanes. *Journal of the American Chemical Society*, 136(33):11570–11573, 2014.
- [51] Bin Xu, Shou-Fei Zhu, Zhi-Chao Zhang, Zhi-Xiang Yu, Yi Ma, and Qi-Lin Zhou. Highly enantioselective s–h bond insertion cooperatively catalyzed by dirhodium complexes and chiral spiro phosphoric acids. *Chemical Science*, 5(4):1442–1448, 2014.
- [52] Shou-Fei Zhu, Yu Fu, Jian-Hua Xie, Bin Liu, Liang Xing, and Qi-Lin Zhou. Synthesis of chiral 4, 4'-disubstituted 1, 1'-spirobiindane-7, 7'-diols and related phosphoramidites: the substituent effect of siphos ligands in rh-catalyzed asymmetric hydrogenation. *Tetrahedron: Asymmetry*, 14(20):3219–3224, 2003.
- [53] Guanghui Wang, Liang Xu, and Pengfei Li. Double n, b-type bidentate boryl ligands enabling a highly active iridium catalyst for c–h borylation. *Journal of the American Chemical Society*, 137(25):8058–8061, 2015.
- [54] Ian Fleming, Rolf Henning, and Howard Plaut. The phenyldimethylsilyl group as a masked form of the hydroxy group. *Journal of the Chemical Society, Chemical Communications*, (1):29–31, 1984.
- [55] H. B. Schlegel G. E. Scuseria M. A. Robb J. R. Cheeseman G. Scalmani V. Barone G. A. Petersson H. Nakatsuji X. Li M. Caricato A. Marenich J. Bloino B. G. Janesko R. Gomperts B. Mennucci H. P. Hratchian J. V. Ortiz A. F. Izmaylov J. L. Sonnenberg D. Williams-Young F. Ding F. Lipparini F. Egidi J. Goings B. Peng A. Petrone T. Henderson D. Ranasinghe V. G. Zakrzewski J. Gao N. Rega G. Zheng W. Liang M. Hada M. Ehara K. Toyota R. Fukuda J. Hasegawa M. Ishida T. Nakajima Y. Honda O. Kitao H. Nakai T. Vreven K. Throssell 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 J. M. Millam M. Klene C. Adamo R. Cammi J. W. Ochterski R. L. Martin K. Morokuma O. Farkas J. B. Foresman M. J. Frisch, G. W. Trucks and D. J. Fox. Gaussian 09 Revision D.01, 2016. Gaussian Inc. Wallingford CT.
- [56] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993.
- [57] Stefan Dapprich, István Komáromi, K. Suzie Byun, Keiji Morokuma, and Michael J Frisch. A new oniom implementation in gaussian98. part i. the calculation of energies, gradients, vibrational frequencies and electric field derivatives 1 dedicated to professor keiji morokuma in celebration of his 65th birthday.1. *Journal of Molecular Structure: THEOCHEM*, 461-462:1–21, 1999.
- [58] Yan Zhao and Donald G. Truhlar. 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. *Theoretical Chemistry Accounts*, 120(1):215–241, May 2008.

- [59] Tobias Risthaus and Stefan Grimme. Benchmarking of london dispersion-accounting density functional theory methods on very large molecular complexes. *Journal of chemical theory and computation*, 9(3):1580–1591, 2013.
- [60] Yi-Pei Li, Joseph Gomes, Shaama Mallikarjun Sharada, Alexis T Bell, and Martin Head-Gordon. Improved force-field parameters for qm/mm simulations of the energies of adsorption for molecules in zeolites and a free rotor correction to the rigid rotor harmonic oscillator model for adsorption enthalpies. *The Journal of Physical Chemistry C*, 119(4):1840–1850, 2015.
- [61] Guilian Luchini, J Alegre-Requena, Ignacio Funes-Ardoiz, and Robert S Paton. Goodvibes: automated thermochemistry for heterogeneous computational chemistry data. *F1000Research*, 9, 2020.

Chapter 3

Analysis of Laboratory-Evolved Flavin-Dependent Halogenases Affords a Computational Model for Predicting Halogenase Site Selectivity

3.1 Author List and Affiliation

Mary C. Andorfer^{3,†}, Declan Evans^{2,†}, Song Yang², Cyndi Qixin He², Anna M. Girlich³, Jaylie Vergara-Coll³, Narayanasami Sukumar⁴, K. N. Houk^{2,*}, Jared C. Lewis^{1,*}

¹Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

²Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, California 90095, United States

³Department of Chemistry, University of Chicago, Chicago, Illinois 60637, United States

⁴NE-CAT and Department of Chemistry and Chemical Biology, Cornell University, Building 436E, Argonne National Laboratory, Argonne, Illinois 60439, United States

[†]These authors contributed equally

^{*}Corresponding authors

This work is a collaboration between the groups of Jared Lewis (Indiana) and K. N. Houk (UCLA), published in *Chem Catalysis* at *Chem Catalysis* **2022**, 2 (10), 2658-2674. Experiments were led by Mary Andorfer under the guidance of Jared Lewis. DFT calculations and MD simulations were led by Declan Evans under the guidance of K. N. Houk.

3.2 Abstract

Flavin-dependent halogenases (FDHs) catalyze selective halogenation of electron-rich aromatic compounds without the need for harsh oxidants required by conventional oxidative halogenation reactions. Predictive models for halogenase site selectivity could greatly improve their utility for chemical synthesis. Toward this end, we analyzed the structures and selectivity of three halogenase variants evolved to halogenate tryptamine with orthogonal selectivity. Crystal structures and reversion mutations revealed key residues involved in altering halogenase selectivity. Density functional theory calculations and molecular dynamics simulations are both consistent with hypohalous acid as the active halogenating species in FDH catalysis. This model was used to

accurately predict the site selectivity of halogenase variants toward different synthetic substrates, providing a valuable tool for implementing halogenases in biocatalysis efforts.

3.3 Introduction

Flavin-dependent halogenases (FDHs) natively catalyze selective halogenation of electron-rich aromatic compounds [1]. The utility of these enzymes stems from their unique mechanism, which involves the reaction of FDH-bound reduced flavin with O₂ to form a peroxyflavin species that reacts with proximally-bound chloride or bromide to generate HOX (X = Cl, Br) [2]. This species migrates through the enzyme to the substrate binding site, where it forms a persistent intermediate thought to be a lysine-derived haloamine [3]. It is not known if this intermediate directly halogenates substrates [3] or serves as a reservoir of X⁺ that can be used to reform HOX upon reaction with water [4]. In either case, halogenation proceeds via electrophilic aromatic substitution (EAS). Although conventional EAS requires harsh oxidants to generate reactive halonium (X⁺) formal equivalents [5], FDHs use atmospheric oxygen and halide salts. Moreover, whereas EAS is typically subject to substrate-controlled site selectivity, substrate binding within an FDH can override substrate-based selectivity (Figure 3.1) [6, 7], enabling catalyst-controlled site selectivity on a range of substrates [8, 9]. The remarkable selectivity of FDHs has made them appealing starting points for enzyme engineering efforts. To assess the feasibility of tuning FDH selectivity, a number of groups have used targeted mutagenesis to alter the site of halogenation [10]. We used directed evolution to engineer variants of the FDH RebH (UniProtKB/Swiss-Prot: Q8KHZ8.1) that halogenate the C7, C6, and C5 positions of tryptamine with > 90% selectivity (0S, 8F, and 10S, Figure 3.1) [11]. We also established that these enzymes demonstrate high conversion and selectivity on a range of other non-native substrates and therefore became interested in developing a general approach to predict FDH selectivity. Docking calculations revealed multiple substrate binding modes, only some of which were consistent with observed selectivity [8]. This inability to reliably predict relevant binding modes indicated that a more sophisticated approach was

required to model FDH selectivity. We reasoned that 0S, 8F, and 10S, which possess only 5-10 mutations relative to one another, would facilitate comparative analysis of FDH selectivity since wild-type (WT) enzymes with analogous selectivity preferences differ by hundreds of residues.¹ Crystal structures and reversion mutations were used to identify mutations responsible for the observed changes in 0S, 8F, and 10S yield and selectivity. To accurately model FDH selectivity, we computationally evaluated the energetic feasibility of both chloramine and HOCl halogenating agents. We then established whether FDH selectivity could be reliably predicted using molecular dynamics (MD) simulations involving energetically-relevant species in analogy to recent studies on native FDH/substrate pairs [12, 13]. This approach provided insight into the mechanisms by which mutations in the RebH scaffold alter binding and selectivity on tryptamine and added support to the case that HOCl is the active halogenating species in FDH catalysis. Moreover, we established that density functional theory (DFT) calculations paired with docking calculations and MD simulations involving this species constitute a framework for accurately predicting FDH site selectivity on non-native substrates, improving the utility of these enzymes for biocatalysis.

3.4 Results

3.4.1 Reversion of Mutations Reveals Residues that Improve Catalysis by Evolved RebH Variants

Relative to RebH, variants 0S, 8F, and 10S possess 1-11 total mutations that increase k_{cat} by 1.2-5.9-fold and significantly increase K_M for 8F and 10S (Table 3.1) [11]. Each of the mutations in 10S and 8F was individually reverted to the WT residue, and the conversion and selectivity of each reversion variant toward tryptamine was evaluated (Figure 3.2).

The simplest RebH variant in the selectivity lineage, 0S, contains a single point mutation (N470S) that gives rise to a six-fold increase in k_{cat} with minimal change in K_M and selectivity for C7 (> 99%). To establish if the beneficial effects of S470 only result from its smaller

size, S470A was introduced into 0S. Although an increase in conversion compared to RebH was observed for this variant, 0S still outperformed S470A by approximately three-fold.

Variant 10S contains six mutations that enable it to halogenate tryptamine with 95% selectivity for C5. Individually reverting these mutations back to the WT residues showed that three of these six mutations are primarily responsible for the selectivity and/or conversion of 10S (Figure 3.2: H52I, C465F, and S470N). Of these three reversions, H52I and C465F are responsible for the observed C5 selectivity. Although the H52I variant has high conversion relative to 10S, the selectivity reverts to primarily C7. The C465F variant displays reduced selectivity for C5 as well as reduced overall conversion. As is observed in 0S, the S470N mutation affects conversion without significantly altering site selectivity. We wondered whether the precise identity of these three residues mattered, or whether similar residues would also confer the observed changes in selectivity and/or conversion. To investigate this possibility, we also made 10S variants H52F, C465S, and S470A (Figure 3.2). H52F led to a modest reduction in C5 selectivity relative to 10S and displayed higher overall conversion. Reduced conversions were observed for the C465S and S470A variants, indicating that C465 and S470 act as more than just a small replacement for the native Phe and Asn residues, respectively.

Variant 8F, which catalyzes 6-chlorination of tryptamine with 90% selectivity, contains 11 mutations. Individually reverting each of these mutations (Figure 3.2) again revealed that not all of the mutations acquired during the evolutionary campaign have a large effect on 8F activity or selectivity. Seven of the 11 reversions were found to affect the selectivity and/or conversion of 8F (Figure 3.2: M52I, P110S, L130S, P448S, L465F, S470N, and Q509R). Residues 52 and 465 were again found to influence selectivity to the greatest extent; however, variant P110S also displayed reduced site selectivity for the C6 position. Several reversions reduced 8F yield (P110S, L130S, P448S, L465F, S470N, and Q509R), but the largest effect was observed for the P448S variant.

3.4.2 Crystal Structures of the 0S-Tryptamine-FAD, 10S-FAD, and 8F-FAD Complexes Provide Insight into the Effects of Key Mutations on Catalysis

The structures of 0S, 8F, and 10S were solved to resolutions of 2.6, 2.25, and 2.55 Å (PDB IDs: 7JU0, 6P00, and 6P2V), respectively. Consistent with previous structures, the asymmetric unit contained two protein molecules. Aligning each structure to that of RebH (PDB ID: 2OA1) [14] revealed RMSD values of 0.283, 0.616, and 0.301 Å² for 0S, 8F, and 10S, respectively, indicating that subtle changes in structure are responsible for the observed changes in selectivity. Occupancy of different ligands typically observed in FDH structures (FAD and halogenation substrate) varied in the structures obtained. In all structures, electron density for the adenine base and phosphate moieties of FAD was observed, but the isoalloxazine ring showed very little density. The occupancy of the loop comprised of residues 449-455, often disordered in previous RebH structures [3, 14], also varied in the structures obtained.

In the structure of 0S, residues 2-528 were included in chain A and residues 2-529 in chain B with no missing regions in between (530 residues for full length enzyme). Density in the composite omit map was observed for the loop consisting of residues 449-455. Though crystals of 0S, 8F, and 10S were soaked in a solution containing tryptamine, occupancy for this substrate was only observed for 0S. This finding is consistent with the much lower K_M for tryptamine of 0S (10.6 μM) relative to 10S and 8F (160 and 1747 μM , respectively). Tryptamine was found to bind in the active site of 0S in a manner analogous to how L-tryptophan binds to RebH (Figure 3.3A). A hydrogen-bonding network between S470, H109, and E357, which is reminiscent of the catalytic triad found in serine hydrolases [15], is also observed in the 0S structure, and this could alter substrate binding and access relative to the native network containing N470 in RebH.

The 10S structure consisted of residues 2-528 in chain A and residues 2-529 in chain B with the loop between residues 451-455 omitted in both chains. Although density is observed within this region, the main chain cannot be reliably modeled within this density. Numerous attempts were

made to soak tryptamine into 10S crystals; however, even with high concentrations of tryptamine (> 10 mM), no clear density for tryptamine was observed within the active site. Overlaying the structures of 10S and RebH containing L-Trp substrate shows a steric clash between H52 in 10S and L-Trp, suggesting that the combination of H52 and C465 must significantly alter substrate binding (Figure 3.3B).

The 8F structure consisted of residues 1-527 in chain A and residues 2-529 in chain B, but the loop between residues 447-456 was omitted in both chains. Seven mutations played a significant role in 8F (Figure 3.3C). Similar to 10S, when the structure of RebH containing L-Trp is overlaid with 8F, a steric clash is observed between N1 and C2 of the indole moiety of the substrate and the sulfur atom of M52 (3.1 and 2.6 Å, respectively). L465 opens the active site, potentially allowing the substrate to rotate or flip to accommodate the M52 mutation. P110 is flanked by two aromatic residues, H109 and F111, which sandwich tryptophan bound in the overlaid RebH structure (Figure 3.3D).

3.4.3 Computational Analysis Suggests that Bound HOCl is the Active Halogenating Species in FDH Catalysis

As previously noted, a lysine-derived chloramine or bound HOCl, both of undetermined protonation state, have been proposed as the active halogenating species during FDH-catalyzed arene halogenation [3, 4]. In either case, aromatic halogenation is believed to occur via EAS involving rate-limiting electrophilic attack to form a Wheland intermediate followed by low-barrier proton loss to form the aromatic product. The lack of a deuterium kinetic isotope effect (KIE) for RebH-catalyzed tryptophan halogenation is consistent with this mechanism [11]. The ΔG° for formation of the Wheland intermediate ($\Delta G_{\text{Whel}}^\circ$) has been found to correlate strongly with ΔG^\ddagger for this reaction [16]. We therefore reasoned that $\Delta G_{\text{Whel}}^\circ$ values calculated by DFT could provide insight into feasibility of different proposed halogenating agents within FDHs. Scheme 1 and Table 2 show the calculated $\Delta G_{\text{Whel}}^\circ$ using several possible halogen donors for EAS at the the 7-position of indole.

Very high $\Delta G_{\text{Whel}}^\circ$ values were calculated for reactions involving neutral methylchloramine

and HOCl (72.4 and 21.2 kcal/mol; Table 2, Entries 1 and 2), making chlorination by these species unlikely. Protonation of either lowers these $\Delta G_{\text{Whel}}^{\circ}$ values considerably (Table 2, Entries 3 and 4) due to the increased electrophilicity of the chlorine. The extremely negative $\Delta G_{\text{rxn}}^{\circ}$ value calculated for H_2OCl^+ calls into question the existence of such a highly reactive species in an enzyme. Instead, it has been suggested that K79 transiently protonates HOCl during the formation of the Wheland intermediate [12], resulting in a slightly negative $\Delta G_{\text{Whel}}^{\circ}$ (Table 2, Entry 5). Taken together, these energies support two possible chlorinating agents: a positively charged lysine-bound chloramine at K79, and HOCl with general acid catalysis by K79 [2, 4, 12].

Because DFT calculations suggested that two species were plausible halogenating intermediates, we next sought to compare these models using molecular dynamics (MD) simulations of tryptamine binding to RebH. Two systems were prepared to represent the two possible reactant complexes determined by DFT: one with the chlorine covalently bound to K79 (K79-Cl, Figure 3.4a), and one with HOCl held close to K79 (K79-HOCl, Figure 3.4b) [13] using restraints in AMBER (see methods). Each simulation was run in pentaplicate, and the distances between the chlorine and substrate carbons were measured (Figure S3.1). Consistently long C-Cl distances are observed with K79-Cl, and these distances vary greatly between replicas. In contrast, much shorter C-Cl distances are observed with K79-HOCl, and these distances were consistent across all replicas. Similar results are observed in the 0S mutant (Figure S3.2) and in previous simulations of tryptophan binding in PrnA and PyrH13. In the current simulations, the chlorine of HOCl remains closer to C7 than either C6 or C5, which is consistent with the experimentally observed selectivity of the enzyme. These simulations suggest that FDH catalysis likely involves halogenation by HOCl rather than a chloramine and that accounting for this distinction could improve predictive models of FDH selectivity.

3.4.4 DFT Calculations and MD Simulations can be used to Model the Selectivity of RebH Variants

With a model for the reaction mechanism in hand, we next sought to investigate how mutations in 0S, 10S, and 8F give rise to the observed site selectivity of these variants. Transition state structures were calculated using DFT for the general acid-catalyzed electrophilic attack at the 5-, 6- and 7-positions of indole (Figure 3.5). These transition states are all structurally similar. A proton is shared between the methylammonium nitrogen and the HOCl oxygen with a nearly linear N–H–O angle. Simultaneously, the chlorine is shared between the HOCl oxygen and indole carbon with a nearly linear O–Cl–C angle. Energetically, TS6 is most favorable with an activation energy of 18.2 kcal/mol with respect to the complex. TS5 was slightly higher in energy (19.0 kcal/mol) and TS7 was higher still (20.1 kcal/mol). The ability of RebH and 0S to chlorinate the least favorable C7 highlights the importance of substrate binding (i.e. catalyst control) for halogenation of indole-containing substrates [10].

Crystal structures of other FDH-tryptophan complexes show that each enzyme binds its substrate with the carbon to be halogenated closest to the active site lysine [2, 3, 14, 17, 18, 19]. RebH is no exception, as C7 of tryptophan is closest to K79 in this structure [3]. The indole ring is held in place by pi stacking with H109 and F111, and the amine extends upward to hydrogen bond with Y454, E461, and F465 (Figure 3.6). The crystal structure of the 0S-tryptamine complex shows that the N470S mutation does not affect this binding orientation, and the two substrates overlay well (Figure 3.6A). The crystal structures of 10S and 8F variants do not contain tryptamine, so, to probe binding within 10S and 8F, tryptamine was docked into the active site of these structures. Docked structures show that tryptamine binds to 10S and 8F in different orientations than that observed in RebH and 0S (Figure 3.6B, C). In these variants, the indole ring flips to hydrogen bond with the backbone of S110 (P110 in 8F), and different carbons are projected towards the catalytic K79. In 10S, C5 is closest to K79 with a N–C5 distance of 4.3 Å; in 8F, C6 is closest to K79 with a N–C6 distance of 3.9 Å. These orientations are consistent with solved crystal structures of the tryptophan-

5 halogenase PyrH and tryptophan-6 halogenases BorH and Thal (Figure S3.3) [17, 18, 19]. Thus, the crystal structure of the 0S-tryptamine complex and structures of tryptamine docked into 10S and 8F show that the evolved variants all follow the established trend that substrates bind with the halogenation site closest to the active site lysine.

As previously noted, a correlation between halogenation site and proximity of that site to K79 in RebH was not observed in docking simulations of more structurally diverse substrates. We therefore sought to determine if MD simulations using HOCl-bound models of FDH variants could provide improved predictive value in this regard. Simulations of RebH in complex with HOCl and tryptamine show that the chlorine remains closer to C7 than either C5 or C6 for most of the simulation and simulations of 0S show similar behavior (Figures S3.1-S3.2, Table S2). Likewise, previously reported simulations of tryptophan binding in PrnA and PyrH in complex with HOCl show that chlorine remains closer to C7 and C5, respectively, in these systems [13]. These results show that FDHs bind their substrates in an orientation that not only places halogenation sites proximal to the active site lysine, but also appears consistent with a near attack conformation (NAC) [20], those conformations that have the substrate and Cl of HOCl within a distance that is likely to lead to a transition state for reaction. Simulations of 8F and 10S were less clear (Figure S3.4); the carbon closest to chlorine changed frequently, and considerably longer C-Cl distances were observed. These observations are consistent with the higher K_M values experimentally measured for these variants. We reasoned that these C-Cl distances, which are significantly longer than those in the calculated transition state structures, are likely not relevant to the observed selectivity.

To precisely define a NAC for halogenation at each site of tryptamine, intrinsic reaction coordinate calculations were performed on the respective calculated transition state structures, and the reactant complexes were optimized by DFT. The resulting complexes show that a C-Cl distance of 3.0-3.1 Å is ideal for NAC formation. Distances this short were rarely if ever observed in MD simulations using conventional force fields, so this threshold was extended to include any distance below 3.6 Å. The free energy of NAC formation (ΔG_N°) in the enzyme can then be estimated by calculating the relative mole fraction of enzyme-NAC relative to enzyme-ground state in each sim-

ulation. Calculated ΔG_N° values provide a quantitative comparison to intrinsic ΔG_{TS}^\ddagger calculated by DFT.

Using this approach, preferential NAC formation can be observed at C7 for RebH and 0S (Table 3 and S3). This site is favored by roughly 2 kcal/mol in these enzymes, which contrasts with the 1.9 kcal/mol the site is disfavored from DFT-calculated ΔG_{TS}^\ddagger values. While preferential NAC formation at C7 is slightly underestimated, the similarity between values obtained by MD simulations and DFT calculations shows that differences in NAC formation can be useful when gauging the enzymes' ability to override the intrinsic site selectivity of the reaction. In the variants 8F and 10S, there is a clear shift away from NAC formation at C7. Instead, NAC formation is roughly equivalent for both C5 and C6 in these enzymes. While C5 NAC formation is slightly underestimated in 10S, this metric provides a useful explanation for why these variants lack selectivity for the original halogenation site and correctly identifies two other sites where NAC formation is preferential.

3.4.5 Predicting Site Selectivity in Non-native Substrates Beyond Tryptamine

Given successful recapitulation of RebH selectivity toward tryptamine using the methods described above, we next sought to evaluate their efficacy toward a subset of the compounds explored in our previous substrate-activity profiling effort [8]. Relative free energies of Wheland intermediate formation ($\Delta\Delta G_{\text{Whel}}^\circ$) were calculated at each sp^2 hybridized carbon to gauge the intrinsic site selectivity for chlorination (Table 4). As previously noted, Wheland intermediate energies have been shown to correlate strongly with transition state barrier heights, while being easier to determine computationally. Comparison of $\Delta\Delta G_{\text{Whel}}^\circ$ on various sites of the indole ring confirm this to be the case (Figure S3.5). For many of the substrates, including **1**, **2**, **4**, **6**, and **7**, chlorination is observed at the most reactive carbon (lowest $\Delta\Delta G_{\text{Whel}}^\circ$), suggesting that the enzyme has little catalyst control over most non-native substrates. This is not always the case (e.g. substrates **3** and **5**), so further modeling of the enzyme is needed to reliably predict chlorination sites.

Docking the substrates into the RebH active site yielded several low-energy poses for each (Fig-

ures S3.6-S3.12). These different binding orientations often projected different halogenation sites towards the catalytic K79, making it difficult to predict the halogenation site based on these structures alone. In every case, docking poses that were consistent with experimentally-determined selectivity were observed, where the carbon to be halogenated is closest to K79. In the docking poses inconsistent with observed selectivity, a carbon predicted to be unreactive by DFT ($\Delta\Delta G_{\text{Whel}}^{\circ} > 15$ kcal/mol) was closest to K79. Although such poses may exist in the enzyme, selectivity on the native substrate, tryptophan, occurs on a site disfavored by 11.2 kcal/mol (C7 vs. C2), and it is unlikely RebH has a higher degree of catalyst control over non-native substrates. Limiting our search to only include reactive poses (where a site with $\Delta\Delta G_{\text{Whel}}^{\circ} < 11$ kcal/mol is closest to K79) the observed halogenation site was always the site closest to K79.

Finally, MD simulations were conducted to quantify catalyst control in the HOCl-bound enzyme (Table 4 and S4, Figure S3.13). For many of the substrates, including **1**, **3**, **6** and **7**, NAC formation was favored at the observed halogenation site. Moreover, in substrate **3**, preferential NAC formation correctly predicts chlorination at C3 over C5, the most reactive site calculated by DFT. Substrate **4** shows similar NAC formation for both C1 and C3; however, differences in intrinsic reactivity make C1 the only possible halogenation site for this substrate. Halogenation could not be correctly predicted with this method for two of the substrates tested. In substrate **2**, NAC formation predicts chlorination at C3, yet no chlorination is experimentally observed at this site. Instead, chlorination is observed at C1, which is calculated to be the most reactive site. For substrate **5**, poor substrate binding was observed, and no NACs were formed at any site. In this case, chlorination is observed at C9, which is less reactive than C7 but closer to K79 in docking calculations.

3.5 Discussion

3.5.1 Crystal Structures, Reversion Mutations, and MD Simulations Provide Molecular Insight into the Altered Selectivity of Evolved FDHs

The crystal structures of 0S, 8F, and 10S, reversion mutations in these enzymes, and a NAC model for selectivity was used to rationalize the effects of mutations identified via directed evolution. The crystal structure of 0S shows that the only mutation in this enzyme, N470S, is located in the active site but does not contact the substrate. Introducing N470A into RebH does not fully account for the increased conversion of 0S, suggesting that the hydroxyl group of serine, not just its small size, is important for the improved function of 0S. In WT RebH, N470 forms a hydrogen bond network with H109 and E357, and this network is maintained but less persistent in the N470S variant (Figure S3.14). Residue 470 is at the start of a large loop (residues 439-468) that regulates substrate access to the active site. We speculate that relaxed hydrogen bonding to S470 facilitates loop movements required for substrate entry, thus increasing k_{cat} 6-fold [21], while permitting native-like hydrogen bonding to maintain a K_M similar to that of WT RebH.

Residues 52 and 465 impart the largest changes in selectivity to both 10S and 8F. Overlaying the structures of 10S and RebH (PDB ID: 2OA1, Figure 3.3) reveals that I52H sterically restricts one region of the active site whereas F465C expands a different section. Docking calculations show that 10S binds the indole moiety in a different orientation than RebH (Figure 3.6B) to avoid a steric clash between tryptamine and H52 if the native binding mode were adopted (Figure 3.3B). Just as in 10S, mutation of residue 52 to methionine within 8F would be expected to favor the flipped substrate binding conformation based on its size relative to the WT isoleucine residue, and this is supported by docking calculations (Figure 3.7).

While both H52 and M52 favor a flipped substrate binding mode, docking calculations suggest that residues 52 and 465 act in concert to rotate the flipped substrate such that C5 and C6 point toward K79 in 10S and 8F, respectively. H52 in 10S is positioned further into the active site

than M52 in 8F due to hydrogen bonding between the indole nitrogen and the backbone carbonyl of E357. This orientation of H52 rotates tryptamine such that C5 is closer to K79 (Figure 3.7). Mutating residue 52 to Phe leads to a variant with reduced selectivity for C5 (Figure 3.2, 85.5% vs 93.7%), perhaps due to its inability to H-bond with E357. The thioether sidechain of M52 in 8F, although bulky enough to cause substrate flipping, could allow the indole ring to rotate slightly closer to residue 52, which would cause C6 to project towards K79 (Figure 3.7).

Reverting C465 and L465 to Phe in 10S and 8F, respectively, also results in lower selectivity and reduced conversion. In 10S, F465C creates more space in the active site, perhaps making room for the bulkier H52 residue in 10S. Variant 10S C465S exhibits similar site selectivity but 8-fold lower conversion relative to 10S, indicating that any improvements resulting from the reduced size and hydrogen bonding capability of C465 cannot be recapitulated even by the relatively conservative S465 variant. In 8F, docking calculations suggest that L465 leads to a steric clash with the primary amine of tryptamine, leading to rotation of tryptamine to achieve C6 selectivity. This repulsion also breaks a hydrogen bond between the backbone carbonyl of residue 465 and the primary amine of tryptamine that is found in both 0S and 10S, which is consistent with the significantly higher K_M of 8F compared with 10S (1747 vs. 160 μM , respectively). The subtle changes in tryptamine positioning by residues 52 and 465 highlight how these residues act together to tune the selectivity of 10S and 8F.

Although several additional residues modulate the yield and selectivity of 8F, reversions P110S and P448S have the largest effects on 8F yield. Residues H109 and F111 contribute to substrate binding via π -stacking, and MD simulations show that P110 rigidifies F111 to enforce this interaction [13] (Figure S3.15) while shifting the backbone carbonyl to better engage the indole N-H of tryptamine in a hydrogen bond to favor the flipped substrate orientation. Residue 448 is in the active site access loop noted above, so P448 could modulate the dynamics of this loop to facilitate substrate entry.

Comparing the effects of critical mutations in 10S and 8F to those present in natural halogenases reveals interesting parallels. First, the three reported 5-tryptophan halogenases possess F52,

[22, 23, 24] which, as noted above, was nearly as effective as H52 in the engineered 5-tryptamine halogenase 10S. O'Connor and coworkers showed that introducing I52F into RebH results in an inactive variant [25]. Introducing I52H both alone and in combination with F465C into RebH also resulted in inactive variants in our hands, indicating that while an aromatic residue at this position can lead to a 5-halogenase, other mutations are required to enable such a change in RebH. Previously reported 6-tryptophan halogenases possess either V52 and greater sequence similarity to 7-tryptophan halogenases [18, 19] or F52 and more sequence similarity to 5-tryptophan halogenases [26, 27, 28]. Our finding that I52M in 8F confers selectivity for the tryptamine 6-position thus represents a unique solution to achieving this selectivity relative to those observed in natural FDHs. Similarly, all known 5- and 6-Trp-FDHs contain proline at the residue corresponding to residue 110 in RebH, but while P110 emerged in the 8F lineage, it is not present in 10S. Finally, all known 5- and 6-Trp-FDHs contain serine at the site corresponding to residue 470 in RebH, and all 7-Trp-FDHs contain asparagine, yet S470 increases the catalytic activity of 0S without changing its selectivity. In short, laboratory evolution of FDHs to alter site selectivity led to mutations that, in some cases, are similar to those that emerged via natural selection [29], but novel solutions to altered selectivity were also obtained.

3.5.2 Analysis of Laboratory-Evolved Enzymes Affords a Computational Model for FDH Selectivity

Both HOCl and an active site lysine-derived haloamine have been proposed as the active halogenating species in FDH catalysis.^{3,4} DFT calculations show halogenation by either of these species, when protonated, is exergonic. Calculated transition state energies show that K79 can activate HOCl by general acid catalysis, and this finding is consistent with mutagenesis experiments that show K79 is essential for activity⁴. Furthermore, the catalytic lysine of many halogenases has been calculated to be much more acidic than free lysine in solution [30, 31]. We calculate pK_a values of 7.1, 7.0, 6.9, and 6.9 for K79 in RebH, 0S, 10S, and 8F, respectively. These low pK_a values result largely from hydrophobic residues, including A50, I52, I82, and F111, that are proximal to

K79. Furthermore, these values suggest that at physiological pH, K79 can readily donate a proton to HOCl to assist in halogenation. MD simulations of WT RebH using K79-HOCl also give much shorter C-Cl distances than K79-Cl, and these distances are consistent with the observed selectivity. Taken together, these data suggest that chlorination by HOCl provides a better model for site selectivity in RebH than a lysine-derived chloramine.

Calculated transition state energies show that chlorination of tryptamine at C6 is most favored and chlorination at C7 is least favored. Nevertheless, RebH and 0S chlorinate C7 with high selectivity, which highlights the importance of substrate binding. Indeed, all tryptophan FDHs crystallized with substrates to date bind their substrates in orientations that project the Csp²-H bond to be halogenated proximal to the active site lysine residue [2, 3, 14, 17, 18, 19]. Although 10S and 8F were not crystallized with substrate, docking calculations show these enzymes also bind tryptamine following this trend. MD simulations using HOCl show that each enzyme preferentially forms near attack conformations for the observed halogenation site. Comparison of $\Delta\Delta G_{\text{N}}^{\circ}$ values determined by MD simulations to $\Delta\Delta G_{\text{T}S}^{\ddagger}$ values calculated by DFT shows that preferential NAC formation accounts for almost all the catalyst control over the reaction.

3.5.3 The FDH Selectivity Model Accurately Predicts Selectivity on Non-Native Substrates

Beyond providing insights into the selectivity of 0S, 8F, and 10S toward tryptamine, computational tools also provided improved ability to predict the selectivity of other non-native FDH-substrate pairs. We previously attempted to use docking calculations alone to model tryptamine binding in 0S, 10S, and 8F11 and to model the binding poses of a panel of substrates in RebH8. In both cases, poses consistent with the observed halogenation selectivity were observed, but so too were poses inconsistent with the observed selectivity and scope.

Calculating the relative energies of Wheland intermediates provided a wealth of predictive ability. Similar metrics, like RegioSQM [32], and HalA [33], have been used previously, and we find that these methods perform comparably (Figure S3.16). For almost every substrate, the site

identified as most reactive was the experimentally observed halogenation site. This is perhaps unsurprising, as these enzymes would not necessarily be expected to exhibit high catalyst control over reactions involving non-native substrates. Docking provided additional insight into substrate-enzyme interactions. As with our previous studies,^{8,11} many docking poses could be identified, and each projected a different halogenation site proximal to K79. Including the $\Delta\Delta G_{\text{Whel}}^{\circ}$ values calculated from DFT allowed us to exclude docking poses that do not represent catalytically active binding conformations (i.e. NACs) and provided a much better predictor for site selectivity. Furthermore, MD simulations show that NAC formation can provide a quantitative comparison between the intrinsic site selectivity of the substrate and the catalyst control of the enzyme.

Density functional theory calculations, docking, and molecular dynamics simulations all provided insight into site selectivity, but no one method was able to reliably predict the halogenation site every time. Instead, the three methods used together provide a more complete picture for substrate and catalyst control over the reaction to identify sites where halogenation is likely. Perfect predictive ability, while ideal, is not completely necessary for a model to provide value. The ability to predict one or a few sites out of many can identify regions of the substrate where chlorination is possible and can be paired with subsequent directed evolution to engineer highly selective catalysts for the desired site.

3.6 Conclusion

Improved models for rationalizing and predicting FDH selectivity toward aromatic substrates would substantially improve the utility of these enzymes for biocatalysis. Our recent finding that FDHs can also catalyze non-native olefin halocyclization suggests that such models would have utility beyond even the native function of these enzymes [34]. We therefore evaluated the energetics of electrophilic aromatic substitution (EAS) by hypohalous acid and chloramine species that have been proposed as active halogenating intermediates in FDH catalysis. The first step of EAS is highly endergonic unless these species are protonated. A low pKa value was calculated for K79,

which is more consistent with its role as a proton donor than chloramine source. Moreover, MD simulations showed that HOCl provided a better model for site selectivity in RebH.

Crystal structures and reversion mutations for three previously evolved RebH variants with orthogonal selectivity were next used to identify mutations that altered variant activity and selectivity. Notably, RebH variants containing only those mutations that led to large changes in selectivity in the evolved variants were inactive, highlighting the importance of permissive mutations, identified via random mutagenesis, for achieving the high selectivities observed. Because these enzymes only possess a handful of mutations, they constitute ideal test cases for predictive models of FDH selectivity. Indeed, docking calculations and MD simulations show that each enzyme binds tryptamine in an orientation that places the carbon to be chlorinated closest the catalytic lysine. This binding orientation allows the enzyme to overcome the intrinsic site selectivity that favors chlorination at the 6-position of tryptamine. Similar calculations also recapitulated the selectivity for several non-native FDH/substrate pairs, which provides a general framework for modeling and predicting FDH selectivity for biocatalysis efforts.

3.7 Methods

3.7.1 Materials

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Oligonucleotides were purchased from Integrated DNA Technologies (San Diego, CA). Plasmids pET-28a/RebF and pET-28a/RebH in BL-21 DE3 E. coli were provided by the Walsh group of Harvard Medical School, Boston, MA [35]. The pLIC-MBP plasmid was provided by the Bottomley group of Monash University, Clayton, Australia [36]. The pGro7 plasmid encoding the groES and groEL chaperone set was purchased from Takara (Otsu, Shiga, Japan). BL21(DE3) E. coli cells were purchased from Invitrogen (Carlsbad, CA). NdeI and HindIII restriction enzymes, T4 DNA polymerase, and Phusion HF polymerase were purchased from New England Biolabs (Ipswich, MA). Luria broth (LB) and Terrific broth (TB) media were purchased from Research Products International (Mt. Prospect, IL). Qiagen Miniprep Kits were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. All genes were confirmed by sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility (900 E. 57th Street, Room 1230H, Chicago, IL 60637). Electroporation was carried out on a Bio-Rad MicroPulser using method Ec2. Ni-nitrilotriacetic acid (Ni-NTA) resin and Pierce[®] BCA Protein Assay Kits were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were followed when using both products. Amicon[®] 30 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C. The glucose dehydrogenase (GDH-105), FAD, and NAD were purchased from Codexis (Redwood City, CA).

3.7.2 General Procedures

Standard molecular cloning procedures were followed [37]. Reactions were analyzed using an Agilent Technologies 6130 LC-MS. RebH protein concentrations were determined by A280 measure-

ments taken on a Tecan Infinite M200 pro microplate reader and extinction coefficients calculated based on amino acid composition (Protein Calculator v3.3, <http://www.scripps.edu/cdputnam/protcalc.html>). MBP-RebF concentrations were measured using the Pierce BCA Protein Assay Kit.

3.7.3 UHPLC/LC-MS Method

Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H₂O/0.1% TFA, solvent B = CH₃CN; 0-10 min, B = 15%; 10-17 min, B = 15-22%; 17-20 min, B = 22-30%; 20-21 min, B = 30%. Absorbance at 280 nm and total ion counts were measured.

3.7.4 Cloning RebH Reversion Variants

All point mutations were introduced via overlap extension technique [38]. The fragment PCR conditions were as follows: 1 ng/ μl parent template, 5x Phusion GC buffer, 0.2 mM dNTPs each, 0.5 μM forward primer, 0.5 μM reverse primer, 0.02 U/ μl Phusion polymerase, and 5% v/v DMSO. Fragments were gel purified. The assembly PCR conditions were as follows: 1:1 ratio of fragments, 5x Phusion GC buffer, 0.2 mM dNTPs each, 0.5 μM forward primer, 0.5 μM reverse primer, 0.02 U/ μl Phusion polymerase, and 5% v/v DMSO. Fragment and assembly PCR were performed in a volume of 50 μL with the following procedure: 98 °C 30 s, (98 °C 20 s, 55 °C 30 s, 72 °C 90 s) for 28 cycles, 72 °C 10 min. The resulting RebH insert was gel purified and digested with the restriction enzymes HindIII (0.33 U/ μl) and NdeI (0.33 U/ μl) in 10x Cutsmart buffer in a final reaction volume of 60 μl . The digestion was conducted at 37 °C for 12-16 hours, after which it was gel purified. This insert was ligated into digested pET-28a (insert:plasmid ratio of 7.5:1) using T4 DNA ligase. Ligations were conducted for 20 hours at 16 °C. Ligations were cleaned with Zymo DNA Cleaning and Concentrating kits and were transformed by electroporation into E. coli BL21(DE3) containing a plasmid encoding the chaperone pGro7.

3.7.5 Expression and Purification of MBP-RebF and RebH

For crystallization experiments, large-scale cultures (750 mL) of RebH variants were grown, expressed, and purified as previously reported [39]. For bioconversions used in selectivity analysis of RebH variants, large-scale cultures (750 mL) of MBP-RebF and smaller cultures (50 mL) of RebH variants were grown, expressed, and purified as previously reported [39, 40].

3.7.6 Specific experimental procedures

Bioconversions with 10S variants

10S variants (0.05 equiv., 25 μ M final concentration) were arrayed into a 96-well microtiter plate in triplicate. MBP-RebF (0.005 equiv., 2.5 μ M final concentration) and glucose dehydrogenase (9 U/mL final concentration) were added as solutions (25 mM HEPES, pH 7.4) to the 10S variants. A solution containing 5-deutero-tryptamine (1 equiv., 0.5 mM final concentration), 6 NAD (0.2 equiv., 100 μ M final concentration), FAD (0.2 equiv., 100 μ M final concentration), NaCl (200 equiv., 100 mM final concentration), phenol (internal standard, 1 equiv., 0.5 mM final concentration) and glucose (40 equiv., 20 mM final concentration) were added via multichannel pipette to simultaneously initiate the reactions (final reaction volume of 75 μ L).

Bioconversions with 8F variants

Bioconversions with 8F variants were conducted similarly to those with 10S variants, except 20 equiv. NaCl (10 mM final concentration) was added to 8F reactions instead of 200 equiv. NaCl (100 mM final concentration). Reactions for all variants were mixed at 650 rpm on top of an Eppendorf air bath and were quenched with 1 volume (75 μ L) methanol after 16 hours. The precipitated protein was removed by centrifugation and the reactions were filtered and analyzed by LC-MS using the method described in General Procedures. Yield and selectivity of all variants were determined following a previously published protocol [40].

3.7.7 Crystallization and structure determination

Following standard Ni-NTA protein purification described previously [39, 40], RebH variants were further purified by gel filtration chromatography using a HiLoad 16/600 Superdex 200 column (GE Healthcare Life Sciences) into a buffer of 20 mM HEPES (pH 7.4) for crystallography. Protein concentration was determined using A280 and extinction coefficients calculated based on amino acid composition (Protein Calculator v3.3, <http://www.scripps.edu/cdputnam/protcalc.html>). Purified protein was concentrated to 8-11 mg/mL, and crystals were grown at 20 °C using the hanging drop vapor diffusion method with a reservoir solution of 1.2-1.3 M Na/K phosphate buffer (pH 6.8) [41, 42]. Single 0S and 10S crystals grew in 1-3 weeks under these conditions. Microcrystalline showers were observed with 8F. 8F microcrystals were used to seed new drops, which produced large, single crystals. Single crystals for 0S, 10S, and 8F were soaked in a solution containing 10 mM tryptamine, 30 mM NaCl, and 5 mM FAD for 5-15 minutes and were subsequently flash frozen in liquid nitrogen following cryoprotection with the reservoir solution supplemented with 16% glycerol. Data were collected at NE-CAT beamline 24-ID-C at the Advanced Photon Source at Argonne National Laboratory using the Pilatus detector. Data were processed using HKL2000 [43]. Phases were determined via molecular replacement using Phaser [44] and wild-type RebH (PDB ID: 2OA1) [45] as the search model. Manual model building was performed in Coot [46] and the structures were refined with PHENIX [47]. The models were validated using composite omit maps for each structure. In the 10S model, residues 451-455 were omitted in both chains because no electron density was observed in this region. In the 8F model, residues 447-456 were omitted in both chains because no electron density was observed in this region. Several 8F structures were solved in attempts to obtain one with this region ordered, since P448 - located distal to the active site, at the end of a flexible, alpha-helical loop - is of particular interest due to its large effect on activity, but all structures contained considerable disorder in this loop region. In all structures, electron density for the adenine base and phosphate moieties of FAD was observed, but the isoalloxazine ring showed very little density. FAD was modeled into the structures based on the

observed density for the base and phosphates, and homology to other FAD-containing structures (PDB ID: 2OA1).

3.7.8 Density functional theory calculations

Conformations of each species were generated using Maestro with the OPLS3 force field. All DFT calculations were performed with Gaussian09 [48] using XSEDE resources [49]. Intrinsic reaction coordinate, geometry optimizations and frequency calculations were performed at the B3LYP-D3/6-31G(d) level of theory [50, 51, 52, 53, 54] using an ultrafine grid and the SMD solvation model [55] for water. Single point energy calculations were performed at the ω B97xD/def2TZVP level of theory [56] using an ultrafine grid and the SMD solvation model for water. Free energies were corrected using the Goodvibes package [57] with Truhlar's rigid-rotor harmonic oscillator treatment [58, 59] for frequencies under 100 wavenumbers.

3.7.9 pKa calculations

All pKa calculations were performed using propka3.1 [60, 61, 62, 63].

3.7.10 Docking calculations

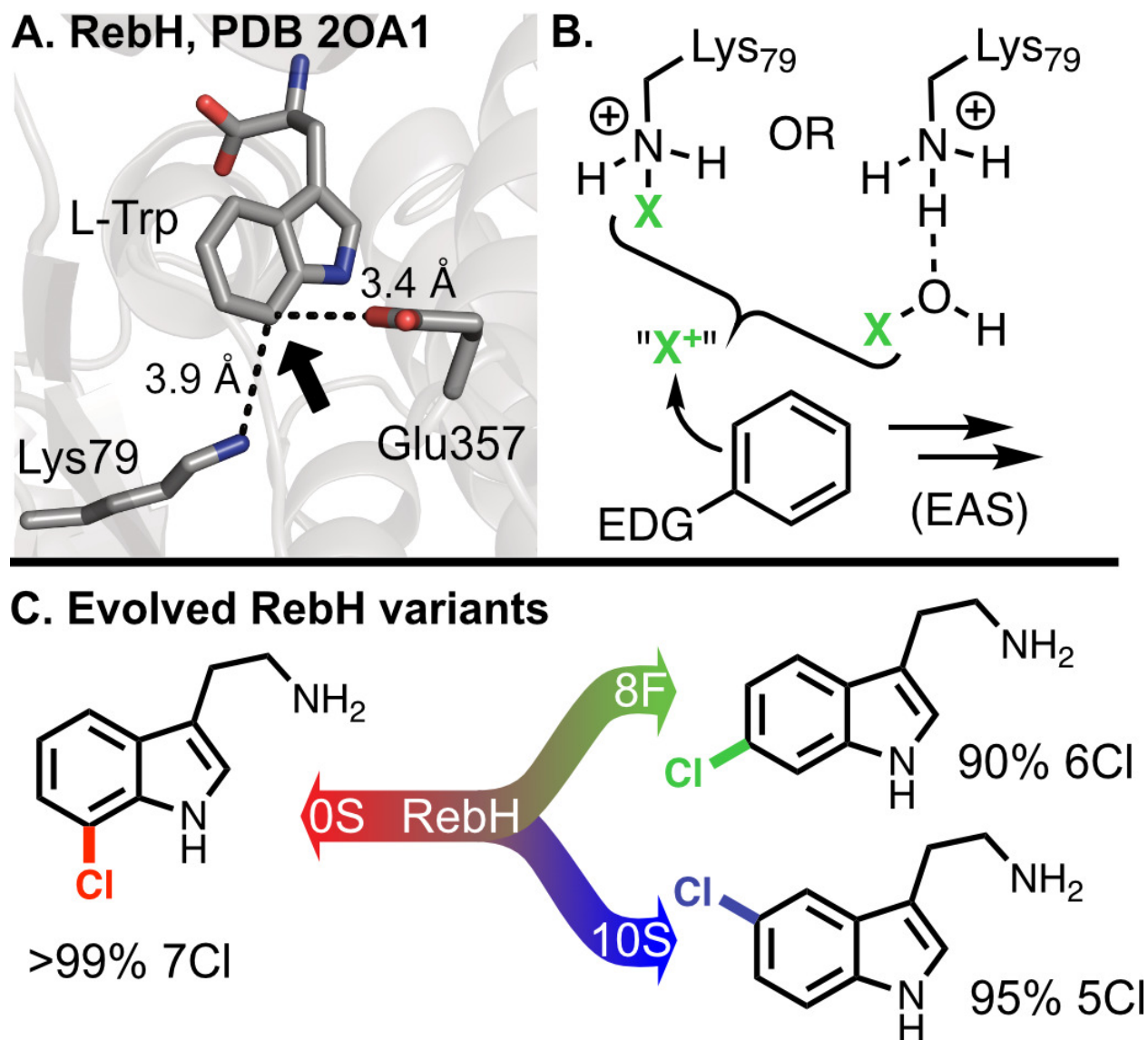
All docking calculations were performed using AutoDock Vina [64]. Crystal structures of each protein (RebH, PDB ID: 2OA1; 10S, PDB ID: 6P2V; and 8F, PDB ID: 6P00) were used as rigid receptors. Chain A was used in every case, and crystallographic Tryptophan was removed from 2OA1 before docking. DFT optimized structures of the substrate were docked into each protein using 20 Å x 20 Å x 20 Å grid boxes centered on the NZ of K79.

3.7.11 Molecular dynamics simulations

Molecular dynamics simulations were performed using the GPU code (pmemd) [65] of the AMBER 16 software package [66]. Parameters for the substrates and noncanonical amino acid K79-Cl

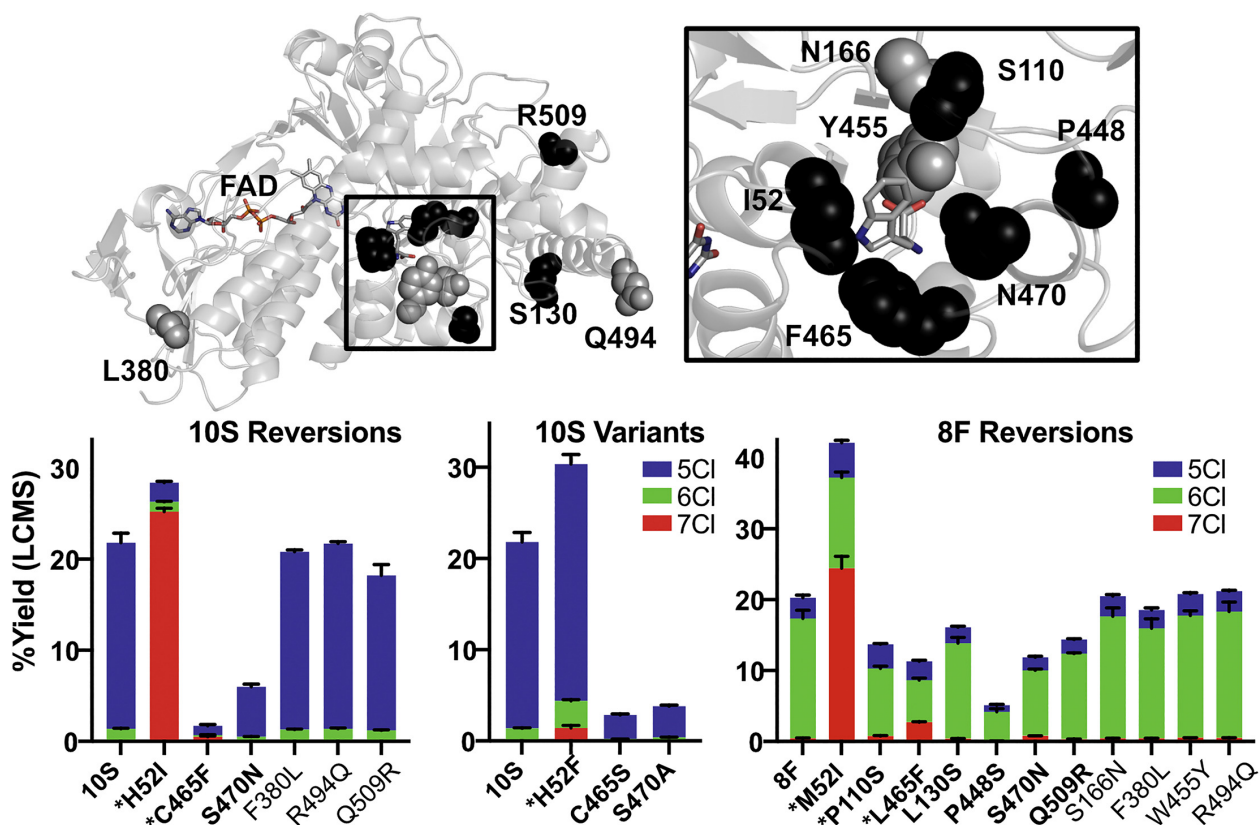
were generated using the antechamber module using the general AMBER force field (GAFF) [67]. Partial charges were set to fit the electrostatic potential generated at the HF/6-31G(d) level of theory by the RESP model [68]. Charges were calculated using the Merz-Singh-Kollman scheme [69, 70] using Gaussian 09. Missing residues in 10S and 8F were built in using the Modeler plugin of UCSF Chimera [71]. Reactant complexes of K79-Cl and K79-HOCl were covalently docked into position using flexible sidechain simulations in AutoDock 4.2.6 [72]. The system was immersed in a pre-equilibrated octahedral box with a 10 Å buffer of TIP3P [73] water molecules using the tleap module. Explicit Na⁺ and Cl⁻ ions were added to neutralize the total charge of the system. All subsequent calculations were performed using the AMBER 14 force field (ff14sb) [74]. Two minimization steps were conducted in serial, each consisting of 2,500 steepest descent steps and 2,500 conjugate gradient steps. In the first equilibration step, the system was heated from 0-300 K over 100 ps using constant-volume and periodic-boundary conditions and a 1 fs time step. The SHAKE algorithm was used to constrain bonds involving hydrogens. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method [75] and an 8 Å cutoff was applied to the Lennard-Jones and electrostatic interactions. A second equilibration was then conducted for 10 ns using a 2 fs time step at constant volume. For runs involving HOCl, DISANG restraints were used to keep the HOCl near K79. The distance between the O of HOCl and NZ of K79 were restrained with distances r1= 1.30, r2= 1.80, r3= 3, r4= 3.50, and force constants rk2=32.0, rk3=32.0. Additionally, the angle between CE of K79, NZ of K79 and O of HOCl were restrained using angles r1= 80, r2= 100, r3= 120, r4= 140, and force constants rk2=25.0, rk3=25.0. These restraints were chosen so that the system would sample conformations only relevant to a reactant complex based on DFT results. Production trajectories were then run in pentaplicate for 500 ns using the same conditions as the second equilibration. Simulations were analyzed using the cpptraj module [76] to extract relevant information, including pdb files for visualization. Plots were generated using in-house python scripts, and pdb files were visualized using open source PyMOL [77].

Figure 3.1: Substrate positioning within flavin-dependent halogenases allows for site-selective halogenation



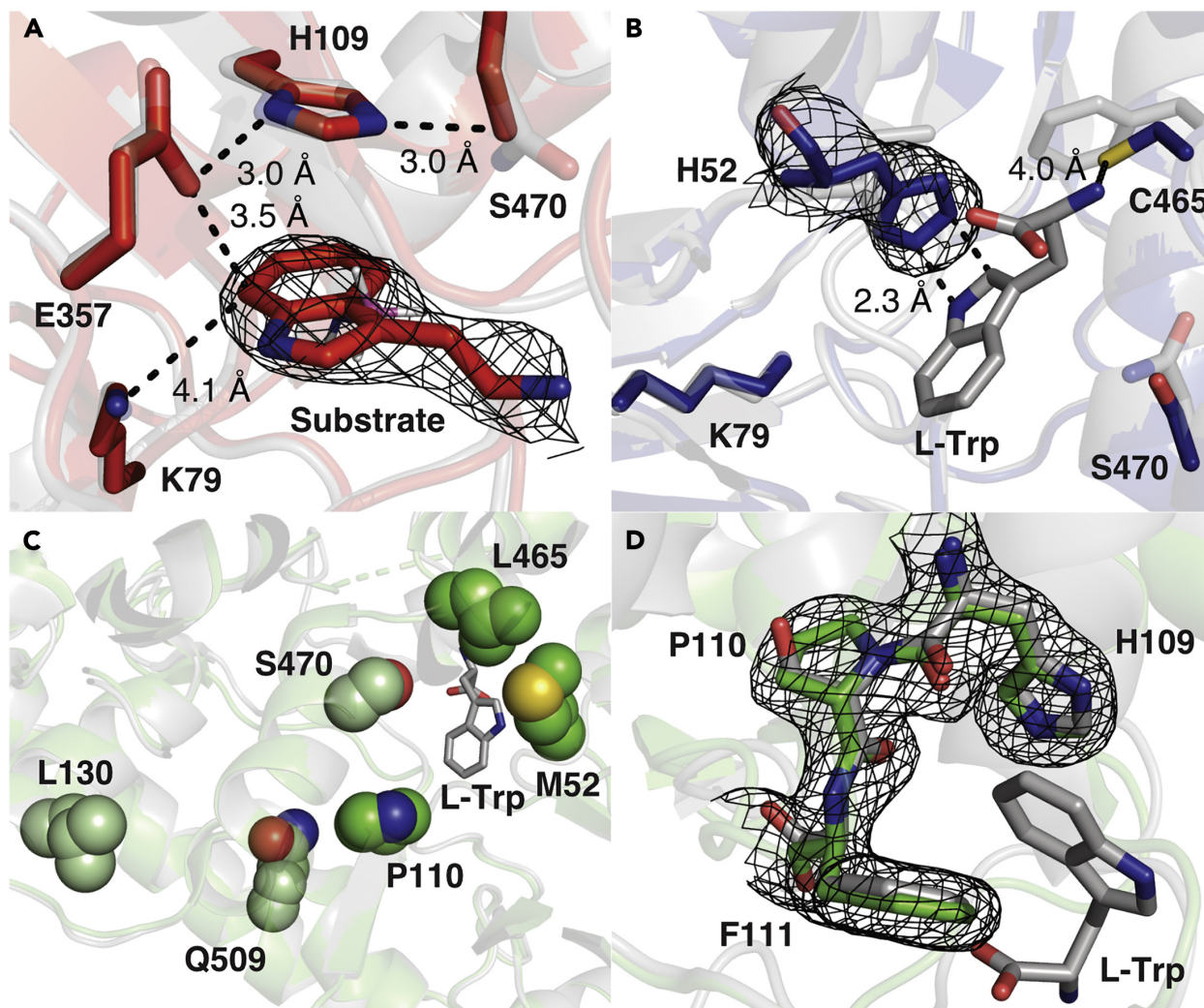
(A) Substrate positioning within FDHs controls selectivity. (B) An electron-rich aromatic ring, shown with an electron-donating group (EDG), undergoes electrophilic aromatic substitution (EAS). Selective EAS catalysis requires an active site lysine residue thought to form a reactive chloramine intermediate or activate HOCl. (C) Directed evolution was used to engineer RebH variants that halogenate tryptamine at C5, C6, or C7 with high site selectivity.

Figure 3.2: Summary of reversion mutations



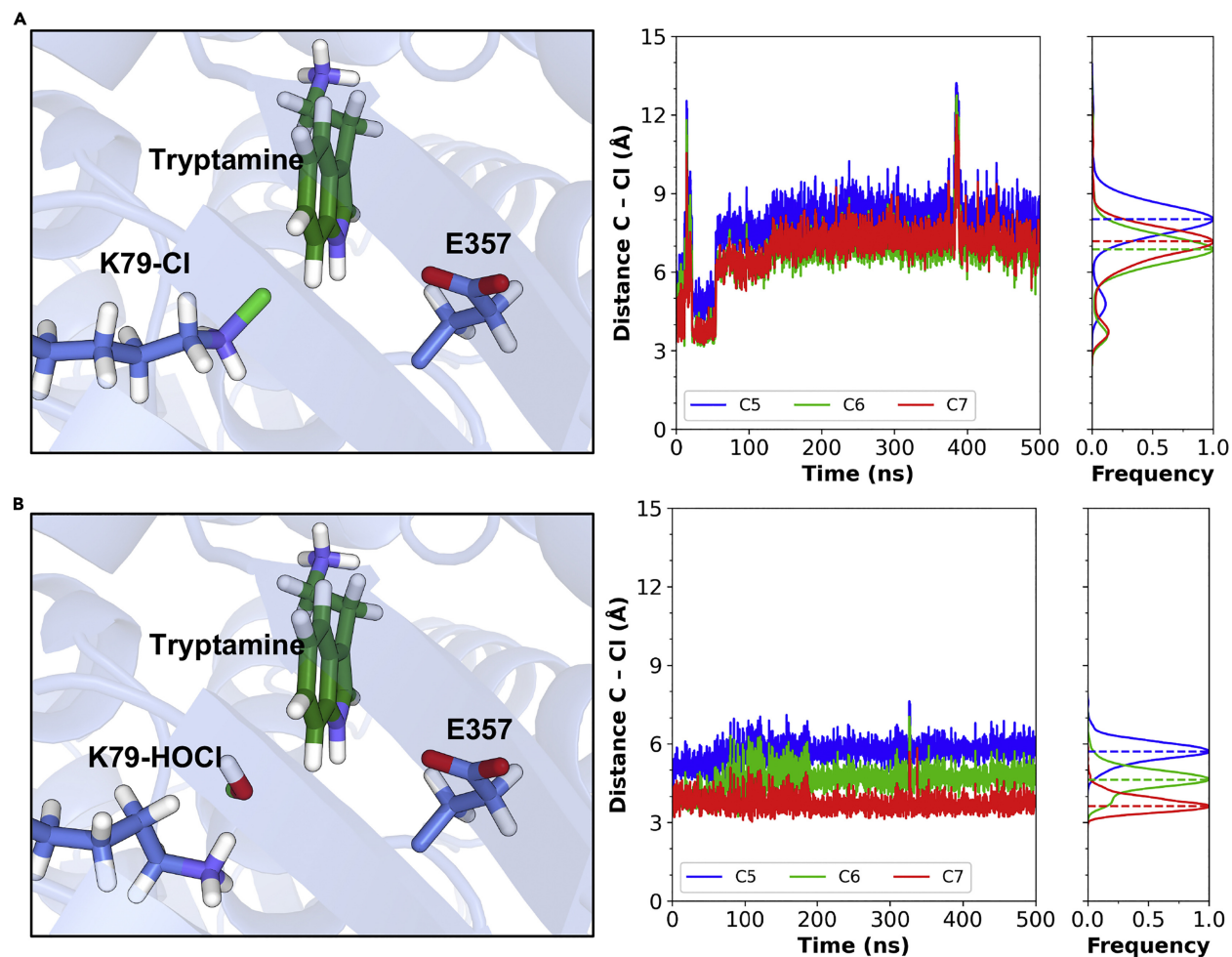
The locations of the 11 sites mutated during the evolutionary campaign to alter site selectivity are shown as spheres within the RebH crystal structure (PDB: 2OA1).¹⁴ Dark gray residues were found to have no effect on yield or selectivity. Black residues were found to affect yield and/or selectivity. FAD and L-Trp are shown as sticks. The magnified inset of the active site is shown in the boxed panel to the right. Graphs indicate yield and selectivity of reversion mutations. Starred residues were found to alter site selectivity. $n = 3$.

Figure 3.3: Comparison of crystal structures between variants



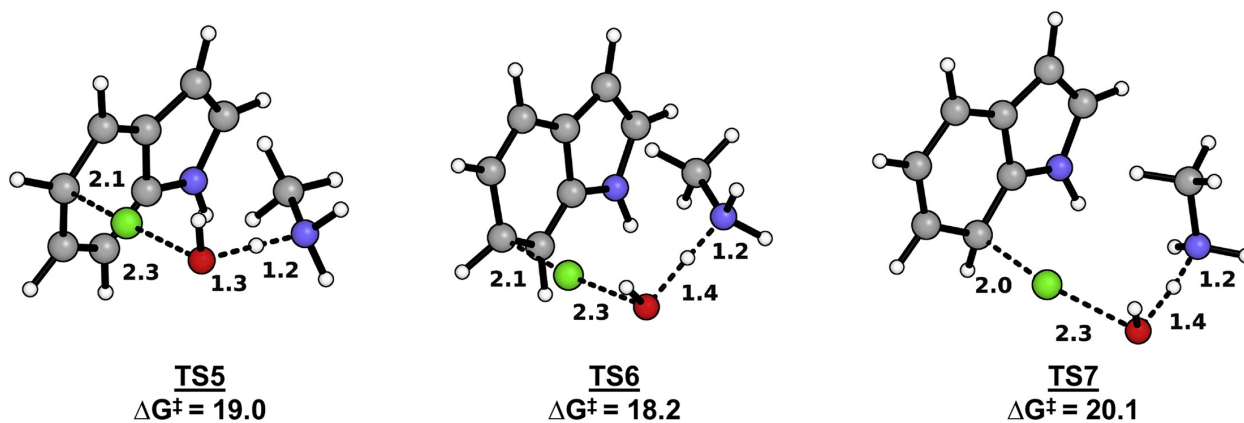
Crystal structures of variants 0S (red), 10S (blue), and 8F (green) overlaid with RebH (gray, PDB: 2OA1). **(A)** Tryptamine bound in the active site of 0S. **(B)** H52 in 10S clashes with L-Trp in RebH. **(C)** Mutations affecting 8F selectivity are shown as darker green spheres. Mutations affecting only 8F yields are shown as light green spheres. **(D)** 8F mutation P110 alters 8F selectivity and yield, perhaps by adding rigidity to neighboring aromatic residues.

Figure 3.4: MD simulations of possible chlorinating species



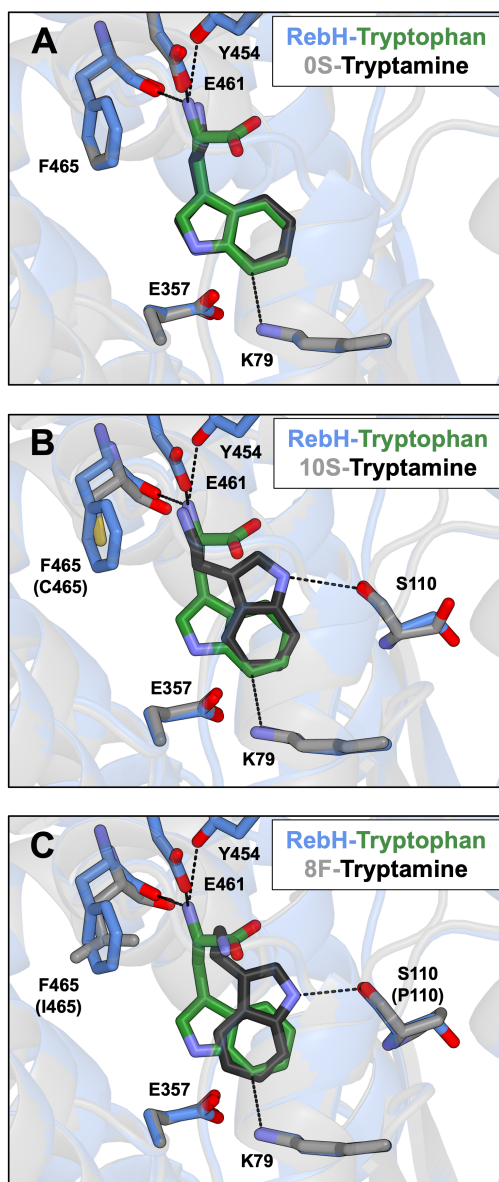
(A) RebH (blue, PDB: 2OA1) with a chlorine covalently bound to K79. Different C-Cl distances are shown (right) over the course of a representative 500-ns simulation. Long distances are observed for each carbon, and these distances vary between replicas. (B) RebH with a HOCl docked into the active site and loosely constrained to K79. Much shorter distances are observed, and these are consistent across replicas. See also Figures S3.1 and S3.2.

Figure 3.5: Transition state structures for the acid-catalyzed chlorination of indole



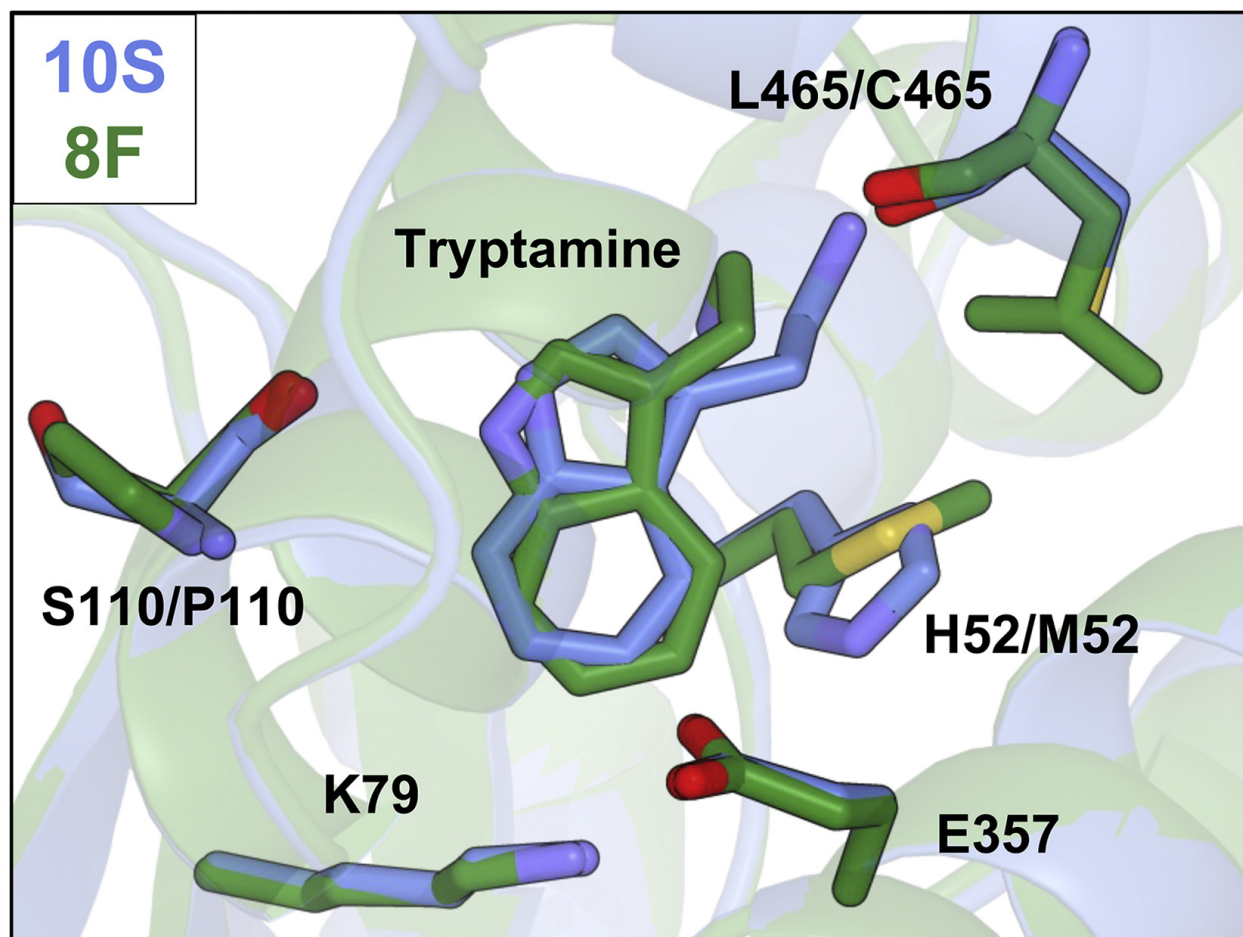
DFT-calculated transition state structures are shown for acid-catalyzed chlorination at the 5-, 6-, and 7-positions of indole. Geometric similarities and energetic differences between structures are highlighted. Chlorination at C6 is most favorable, while C7 is least favorable, according to the calculated activation energy.

Figure 3.6: Substrate binding poses



Substrate binding poses obtained from crystal structures and docking calculations are shown overlaid with the crystal structure of RebH bound to tryptophan (2OA1). Overlays of (A) 0S, (B) 10S, and (C) 8F binding poses.

Figure 3.7: Comparison of substrate binding in variants



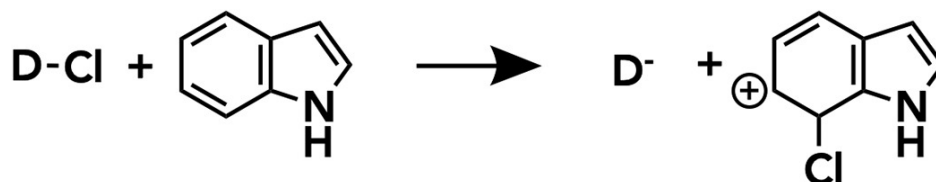
Docked binding orientations in 10S (blue) and 8F (green). Tryptamine positioning is affected by residue 52. H52 in 10S projects C5 toward K79, and M52 in 8F projects C6 toward K79.

Table 3.1: Kinetic parameters of RebH and variants

Enzyme	K_m (μM)	k_{cat} ($\text{min})^{-1}$	k_{cat}/K_m ($\text{min } \mu\text{M})^{-1}$	No. of Mutations
RebH	9	0.023	2.6×10^{-3}	0
0S	10.6	0.135	1.3×10^{-2}	1
8F	1747	0.037	2.1×10^{-5}	11
8F	160	0.028	1.8×10^{-4}	6

Data adapted from Andorfer et al. [11]

Table 3.2: Thermodynamic parameters for electrophilic attack of different Cl⁺ donors on indole



Scheme 1. Electrophilic attack of different halonium donors on the 7-position of indole

Entry	Donor-Cl	Donor ⁻	$\Delta G_{\text{Whel}}^{\circ}$ (kcal/mol)
1	CH ₃ NHCl	CH ₃ NH ⁻	72.4
2	HOCl	HO ⁻	21.2
3	CH ₃ NH ₂ Cl ⁺	CH ₃ NH ₂	1.7
4	H ₂ OCl ⁺	H ₂ O	-57.8
5	HOCl + CH ₃ NH ₃ ⁺	H ₂ O + CH ₃ NH ₂	-3.3

ω B97X-D/def2TZVP-SMD(water)//B3LYP-D3/6-31G(d)-SMD(water)

Table 3.3: Substrate (ΔG_{TS}^\ddagger) vs catalyst (ΔG_N°) control over site selectivity of tryptamine in RebH and variants

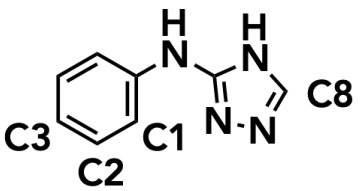
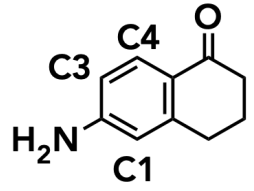
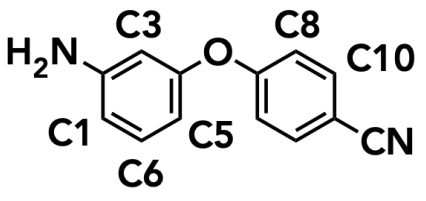
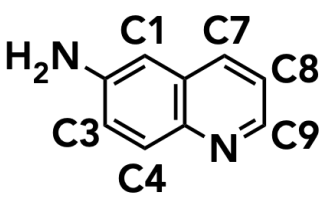
FDH	C _n	$\Delta G_{TS}^{\ddagger a}$ (kcal/mol)	$\Delta G_N^{\circ b}$ (kcal/mol)	$\Delta \Delta G_{TS}^{\ddagger}$ (kcal/mol)	$\Delta \Delta G_N^\circ$ (kcal/mol)	Obs. ^c (%)
RebH	5	19.0	–	0.8	–	0
	6	18.2	2.2	0.0	1.9	0
	7	20.1	0.3	1.9	0.0	99
0S	5	19.0	4.2	0.8	3.5	0
	6	18.2	2.8	0.0	2.1	0
	7	20.1	0.7	1.9	0.0	99
8F	5	19.0	2.2	0.8	0.0	11
	6	18.2	2.3	0.0	0.1	89
	7	20.1	4.0	1.9	1.8	0
10S	5	19.0	4.2	0.8	0.0	94
	6	18.2	4.2	0.0	0.1	6
	7	20.1	–	1.9	–	0

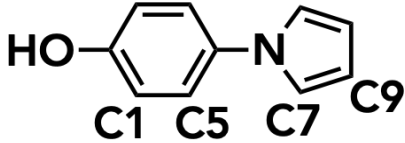
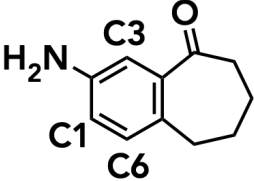
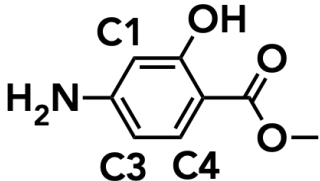
^a Transition state free energy barrier for chlorination in solution calculated by DFT.

^b Free energy of NAC formation in enzyme estimated by MD simulations. Empty values indicate no NAC is formed at this site.

^c Observed (Obs.) percentage of different isomers as determined by ¹H NMR spectroscopy.

Table 3.4: Substrate ($\Delta\Delta G_{\text{whel}}$) vs catalyst ($\Delta\Delta G_{\text{N}}^{\circ}$) control over site selectivity of non-native substrates in RebH

Entry	Compound	C _n	$\Delta\Delta G_{\text{whel}}^{\text{a}}$ (kcal/mol)	$\Delta\Delta G_{\text{N}}^{\text{ob}}$ (kcal/mol)	Obs. (%)
1		1 2 3	5.8 20.9 0.0	– 2.4 0.0	0 0 100
2		1 3 4	0.0 0.4 29.1	– 0.0 4.1	100 0 0
3		1 3 5 6 8 10	0.0 – – 1.6 – –	– 2.4 0.0 0.0 0.0 0.0	100 0 0 0 0 0
4		1 3 4 7 8 9	0.0 18.7 26.6 20.6 29.7 15.2	0.3 0.0 2.3 – – –	100 0 0 0 0 0

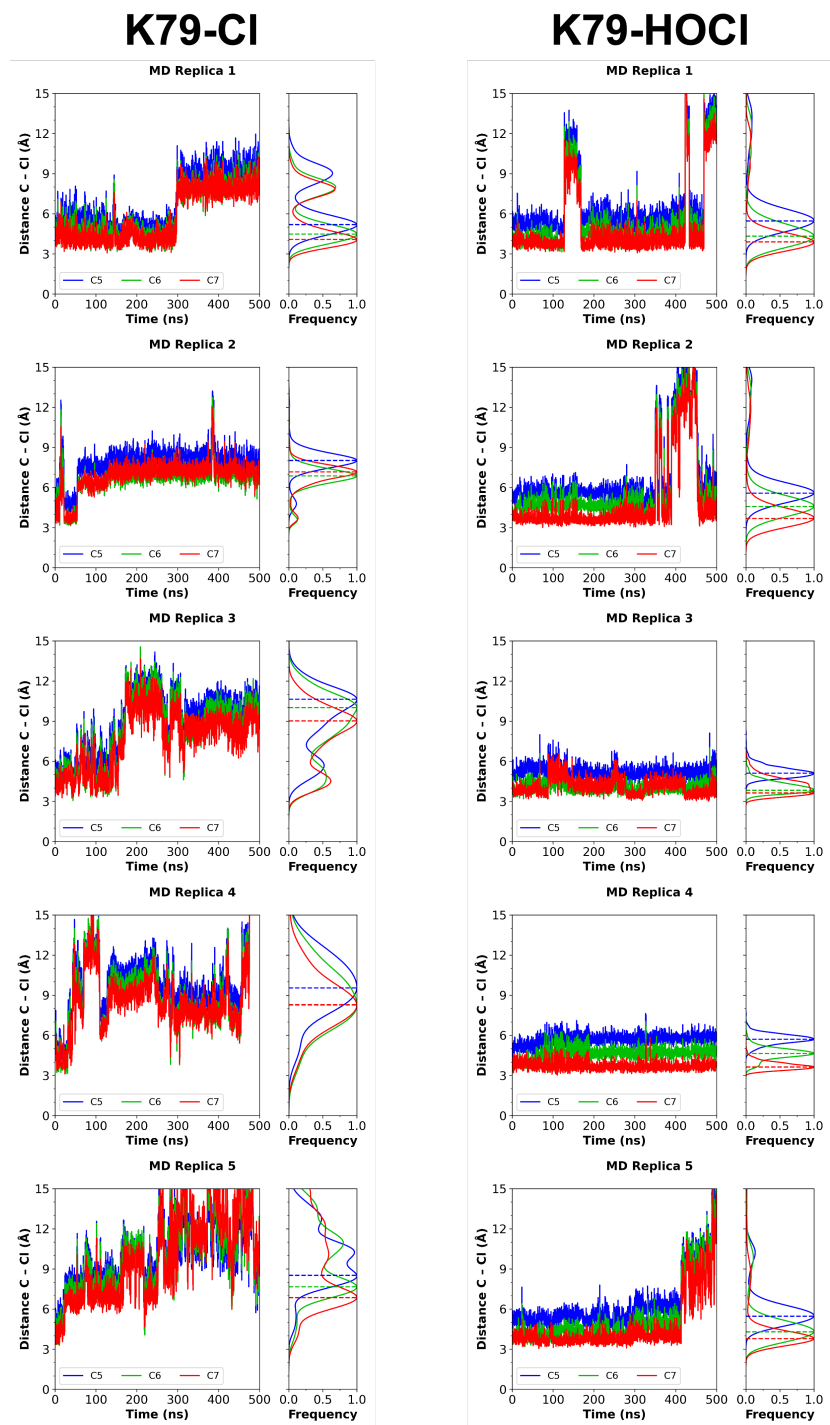
5		1 5 7 9	25.2 26.0 0.0 6.2	– – – –	0 0 0 100
6		1 3 6	0.0 0.7 20.8	0.0 – 2.5	93 7 0
7		1 3 4	1.7 0.0 35.4	– 0.0 –	6 94 0

^a Relative free energy of Wheland intermediate formation at the given site calculated by DFT.

^b Relative free energy of NAC formation in RebH estimated by MD simulations. Empty values indicate that no NAC is formed at this site.

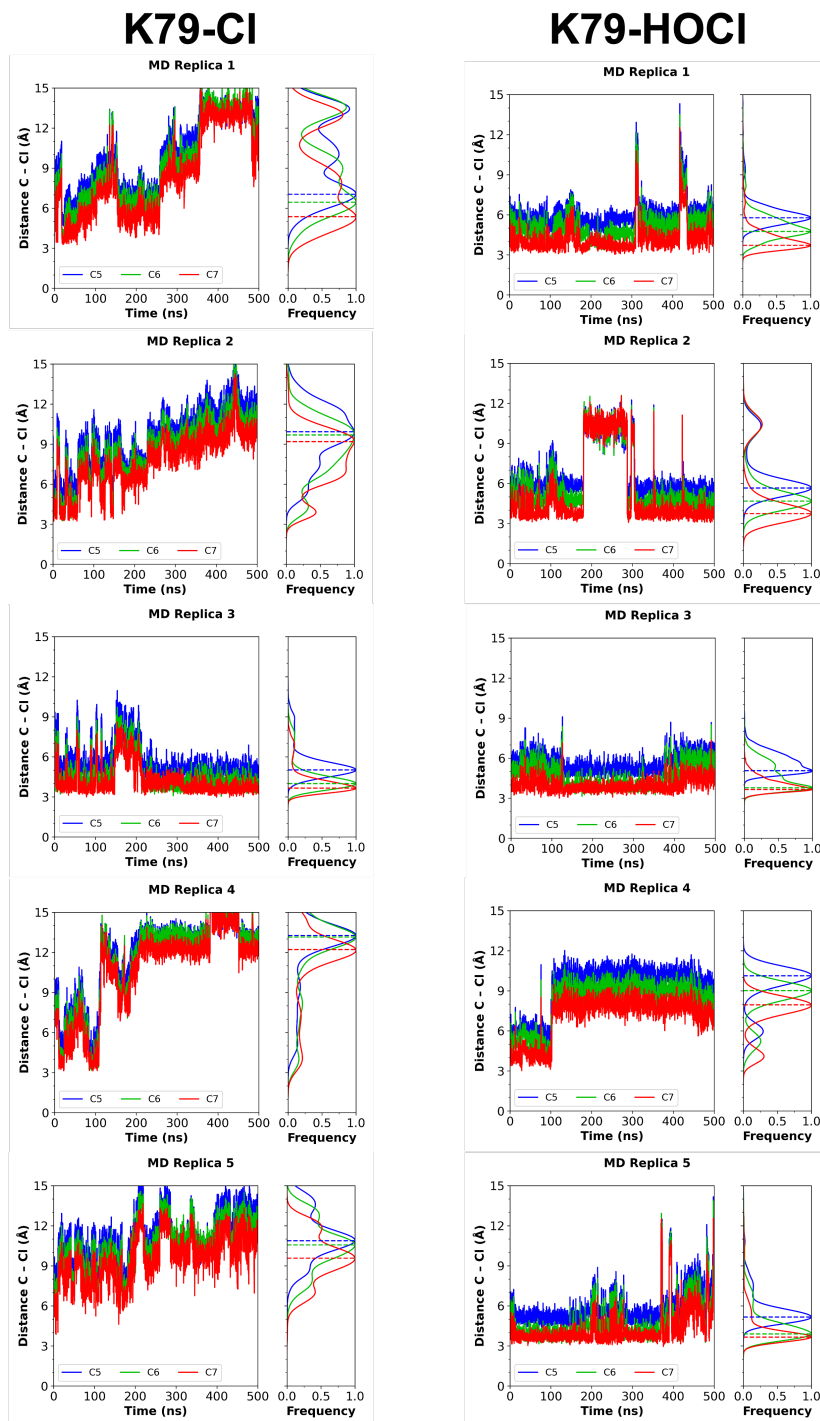
^c Observed (Obs.) percentage of different isomers as determined by ¹H NMR spectroscopy.

Figure S3.1: Molecular dynamics simulations of RebH



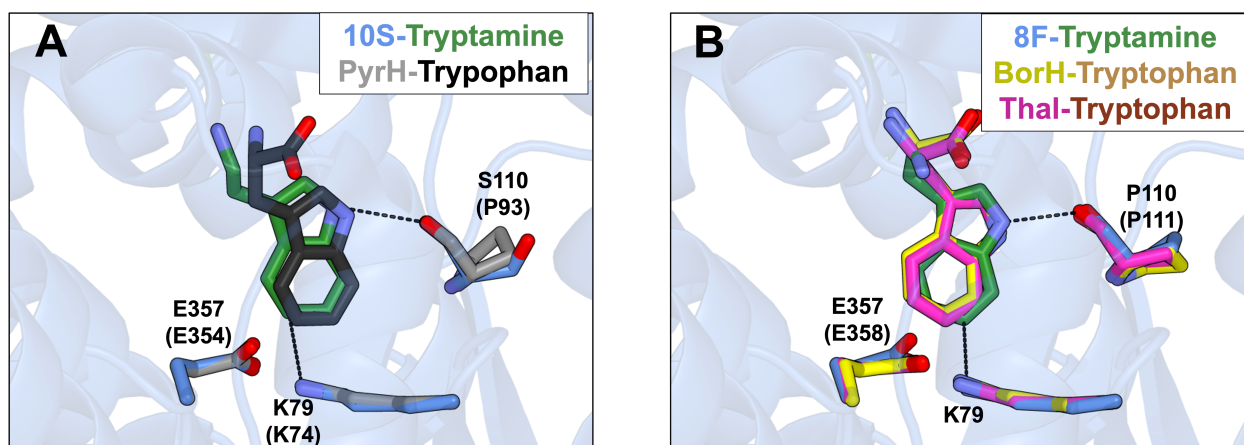
C-Cl distances from molecular dynamics comparing K79-CI and K79-HOCl in wild-type RebH. Each plot represents one replica of a MD simulation using the described condition.

Figure S3.2: Molecular dynamics simulations of 0S



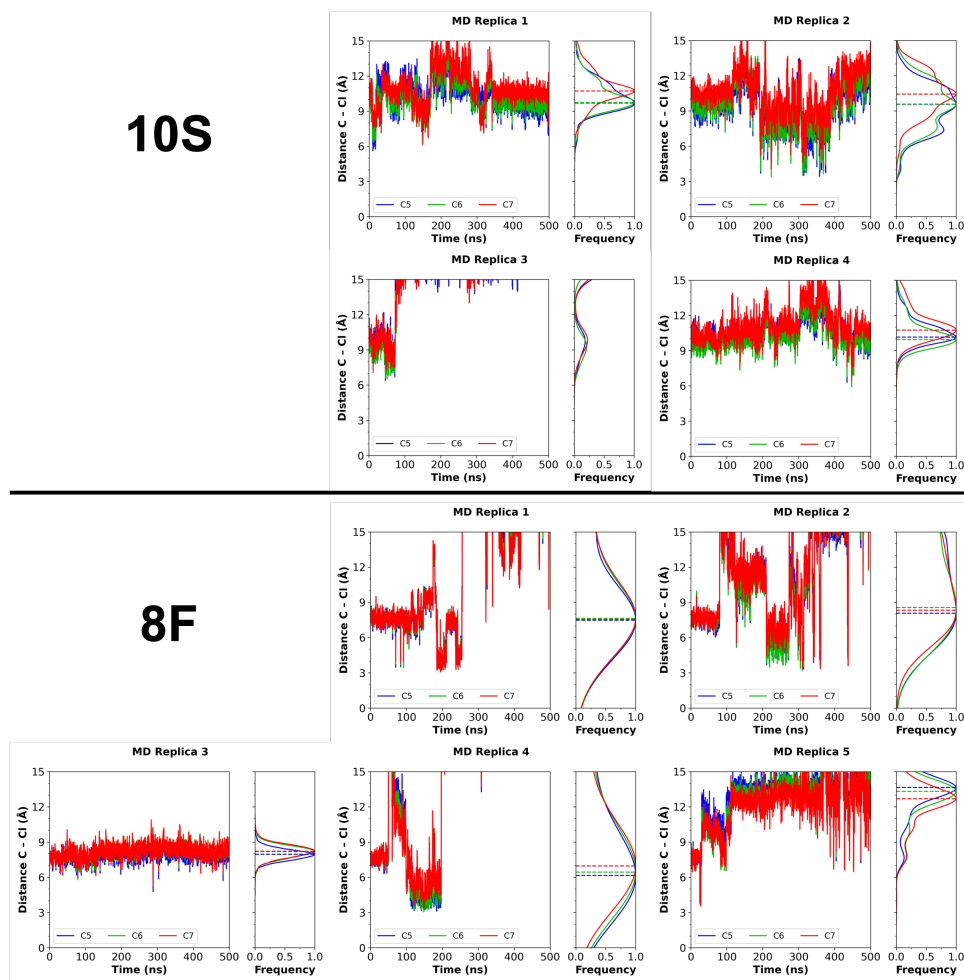
C-Cl distances from molecular dynamics comparing K79-CI and K79-HOCl in 0S. Each plot represents one replica of a MD simulation using the described condition.

Figure S3.3: Comparison of substrate binding poses in variants with other natural halogenases



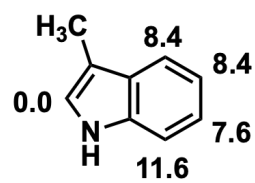
(A) Overlay of 5-selective enzymes 10S and PyrH (2WEU). (B) Overlay of 6-selective enzymes 8F, BorH (6UL2), and Thal (6H44)

Figure S3.4: Molecular dynamics simulations of variants 10S and 8F



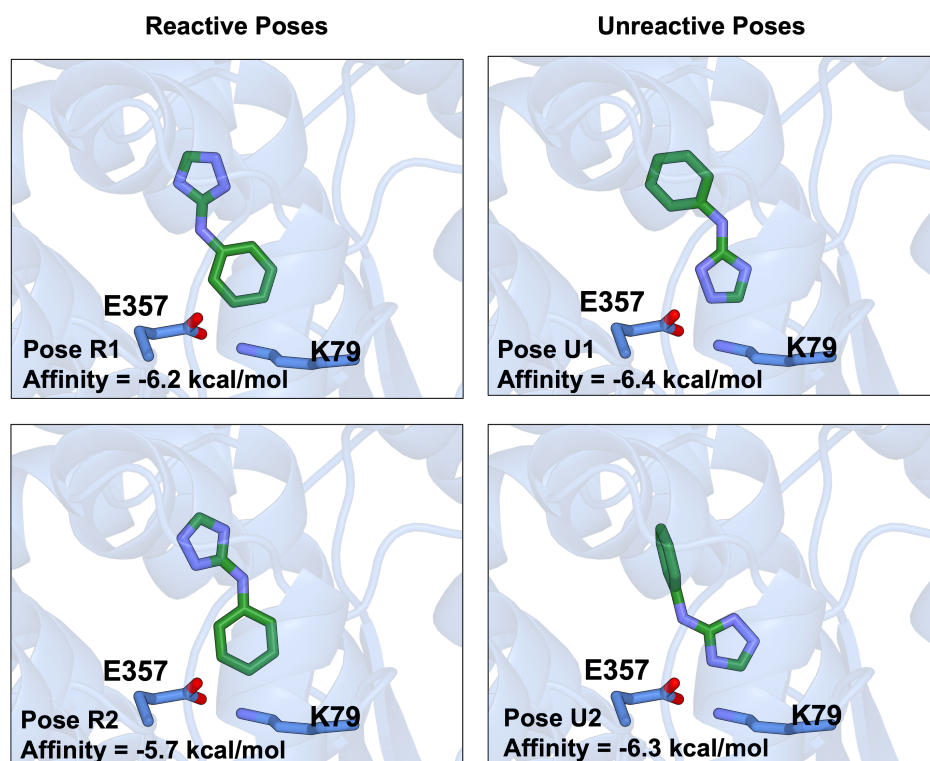
Each plot represents one replica of a MD simulation using K79-HOCl as a chlorine source.

Figure S3.5: Calculated $\Delta\Delta G_{\text{whel}}$ for different sites on the indole ring



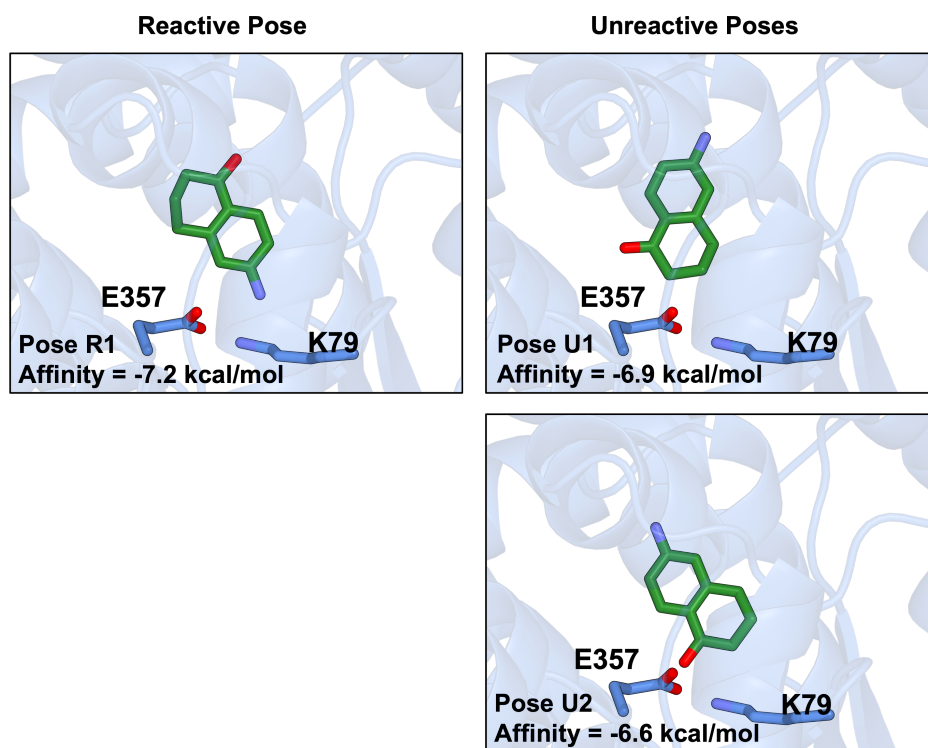
ω B97X-D/def2TZVP-SMD(water)//B3LYP-D3/6-31G(d)-SMD(water)

Figure S3.6: Calculated docking poses for substrate 1



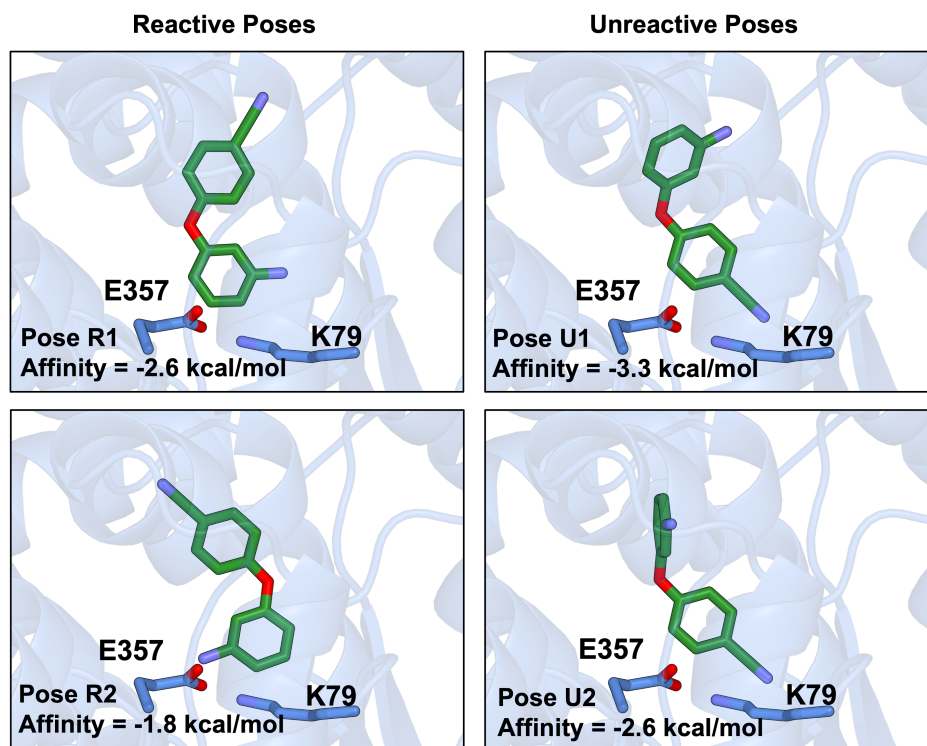
Affinity energies calculated by AutoDock Vina are shown with each docking pose

Figure S3.7: Calculated docking poses for substrate 2



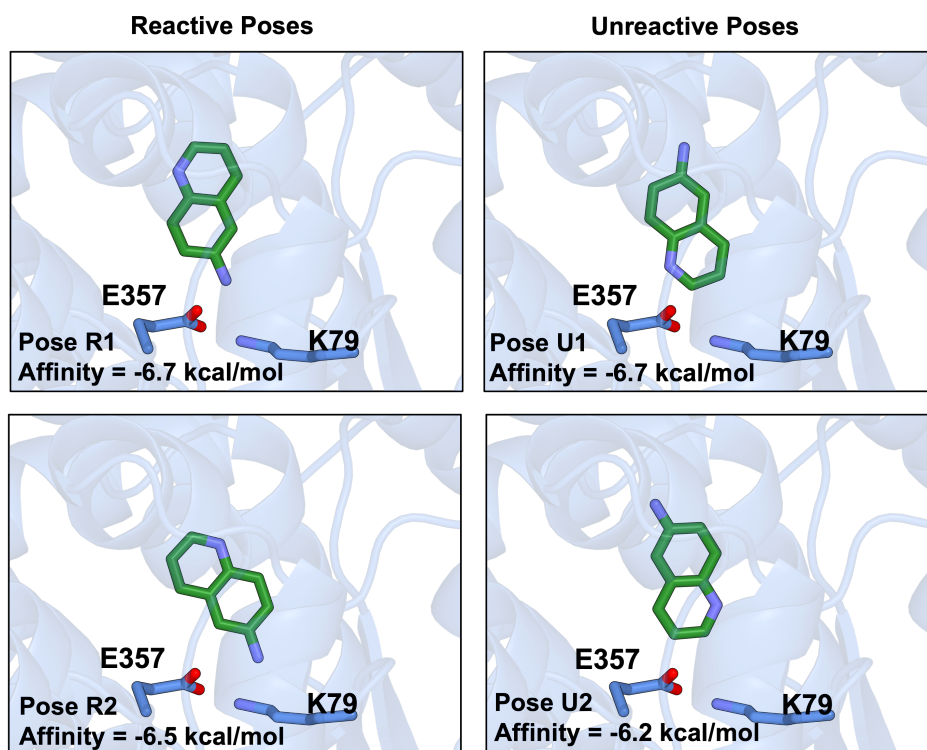
Affinity energies calculated by AutoDock Vina are shown with each docking pose

Figure S3.8: Calculated docking poses for substrate 3



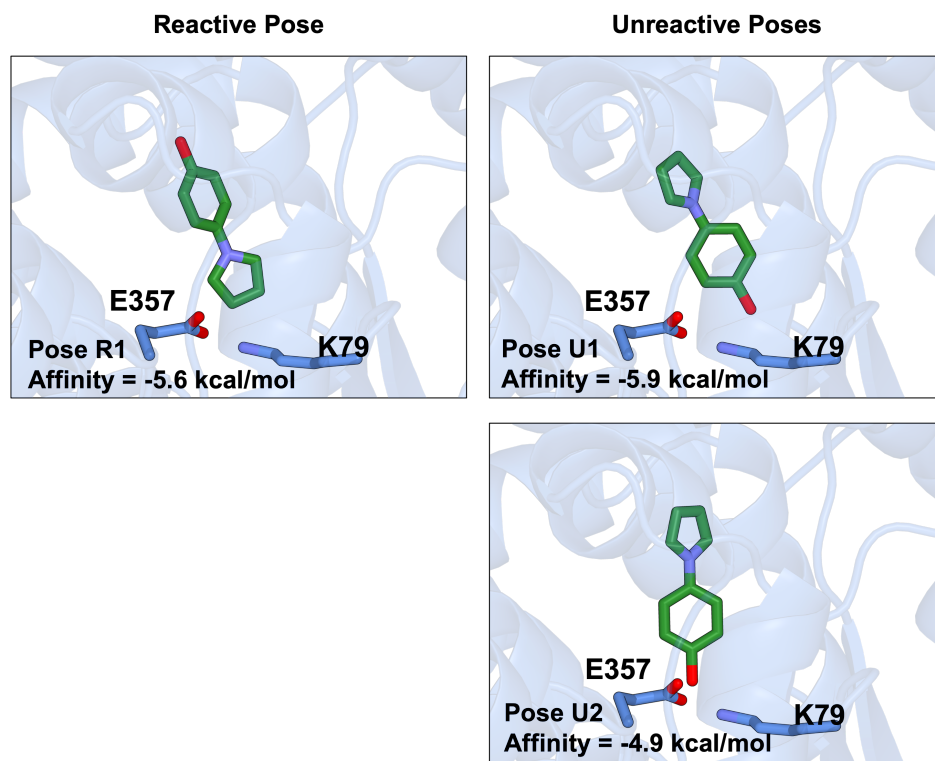
Affinity energies calculated by AutoDock Vina are shown with each docking pose

Figure S3.9: Calculated docking poses for substrate 4



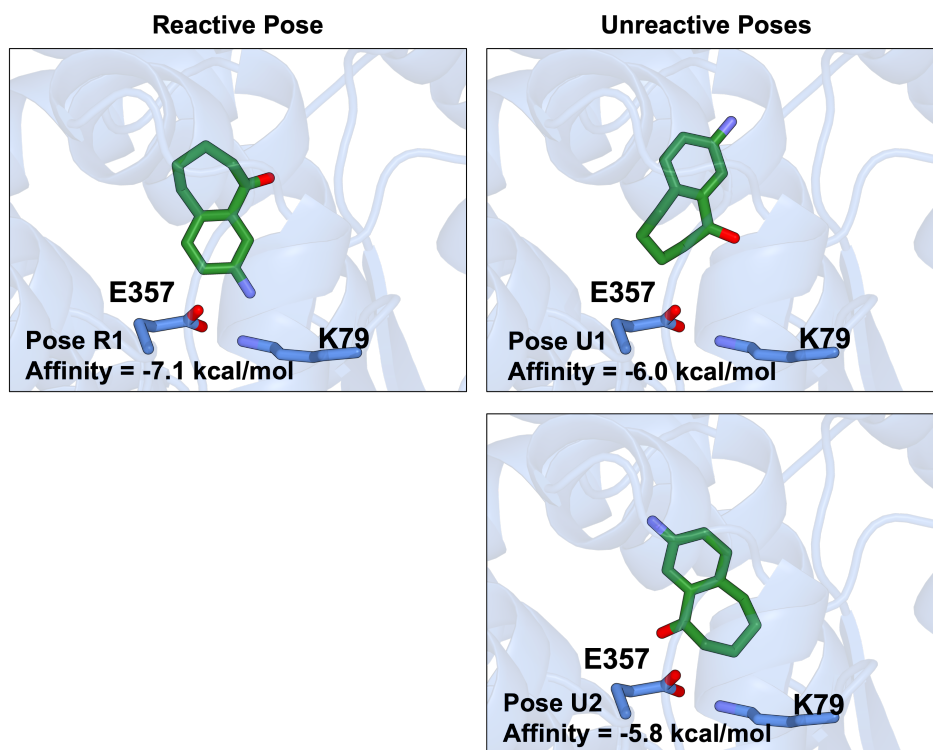
Affinity energies calculated by AutoDock Vina are shown with each docking pose

Figure S3.10: Calculated docking poses for substrate 5



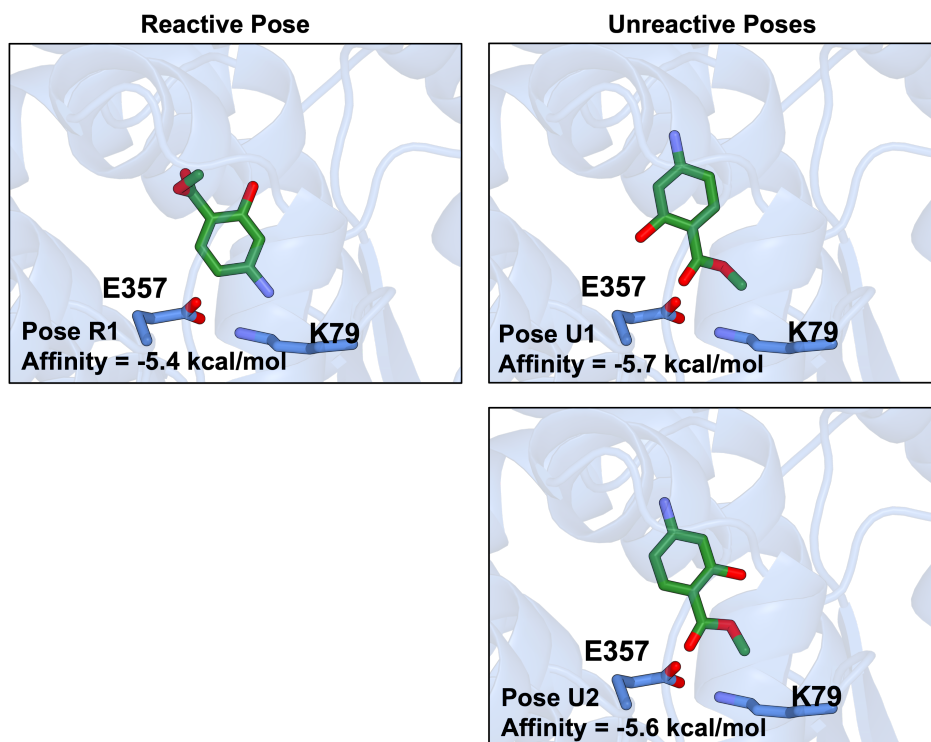
Affinity energies calculated by AutoDock Vina are shown with each docking pose

Figure S3.11: Calculated docking poses for substrate 6



Affinity energies calculated by AutoDock Vina are shown with each docking pose

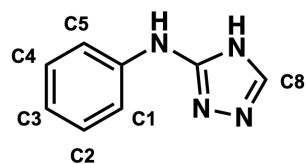
Figure S3.12: Calculated docking poses for substrate 7



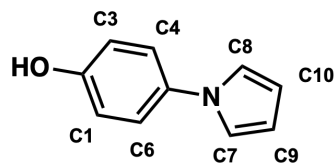
Affinity energies calculated by AutoDock Vina are shown with each docking pose

Figure S3.13: Full carbon numbering for each nonnative substrate used in this study

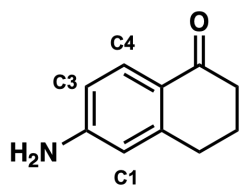
1



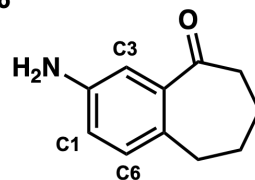
5



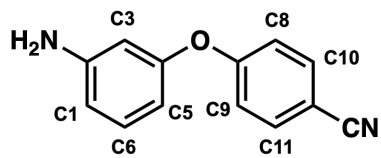
2



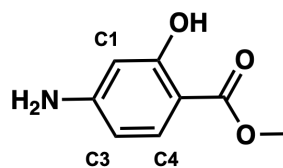
6



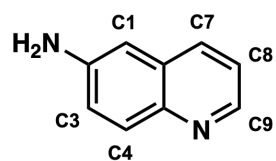
3



7

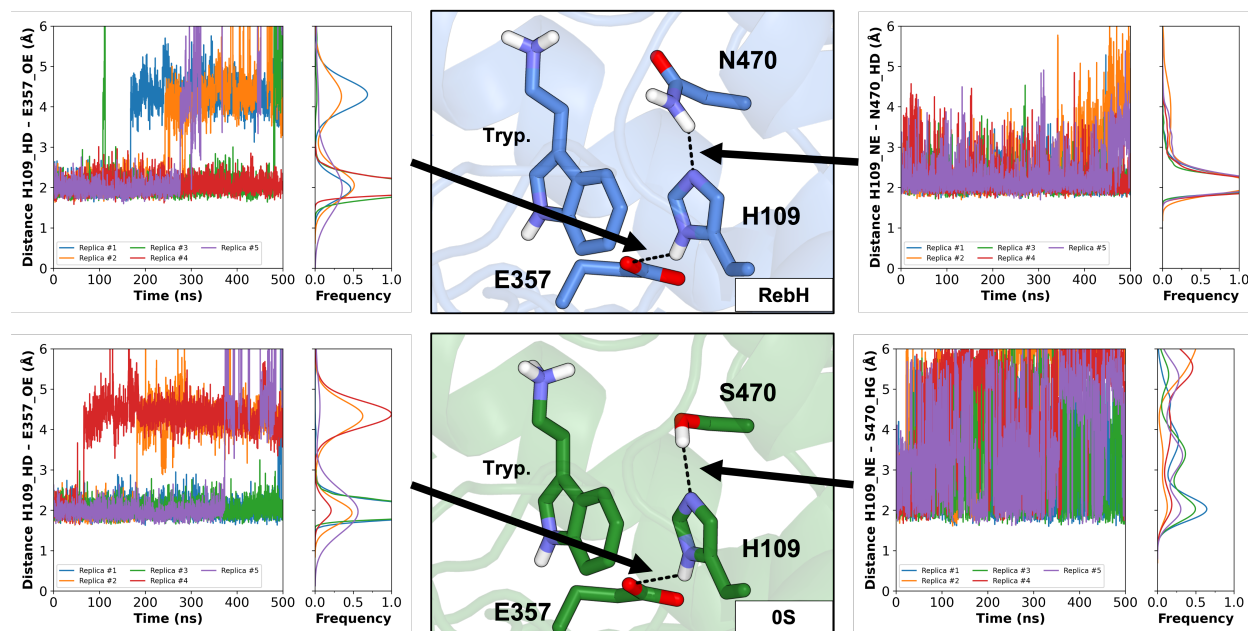


4



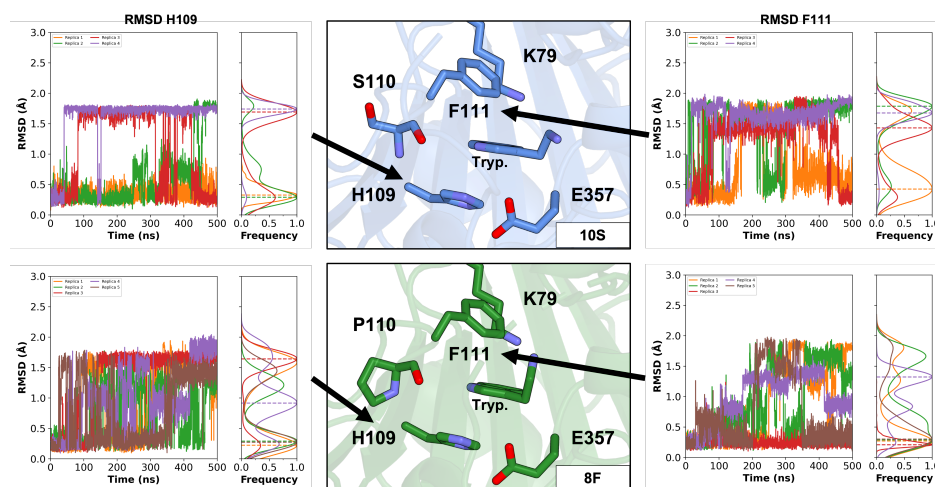
Full carbon numbering for each nonnative substrate used in this study

Figure S3.14: Dynamics of a hydrogen bond triad



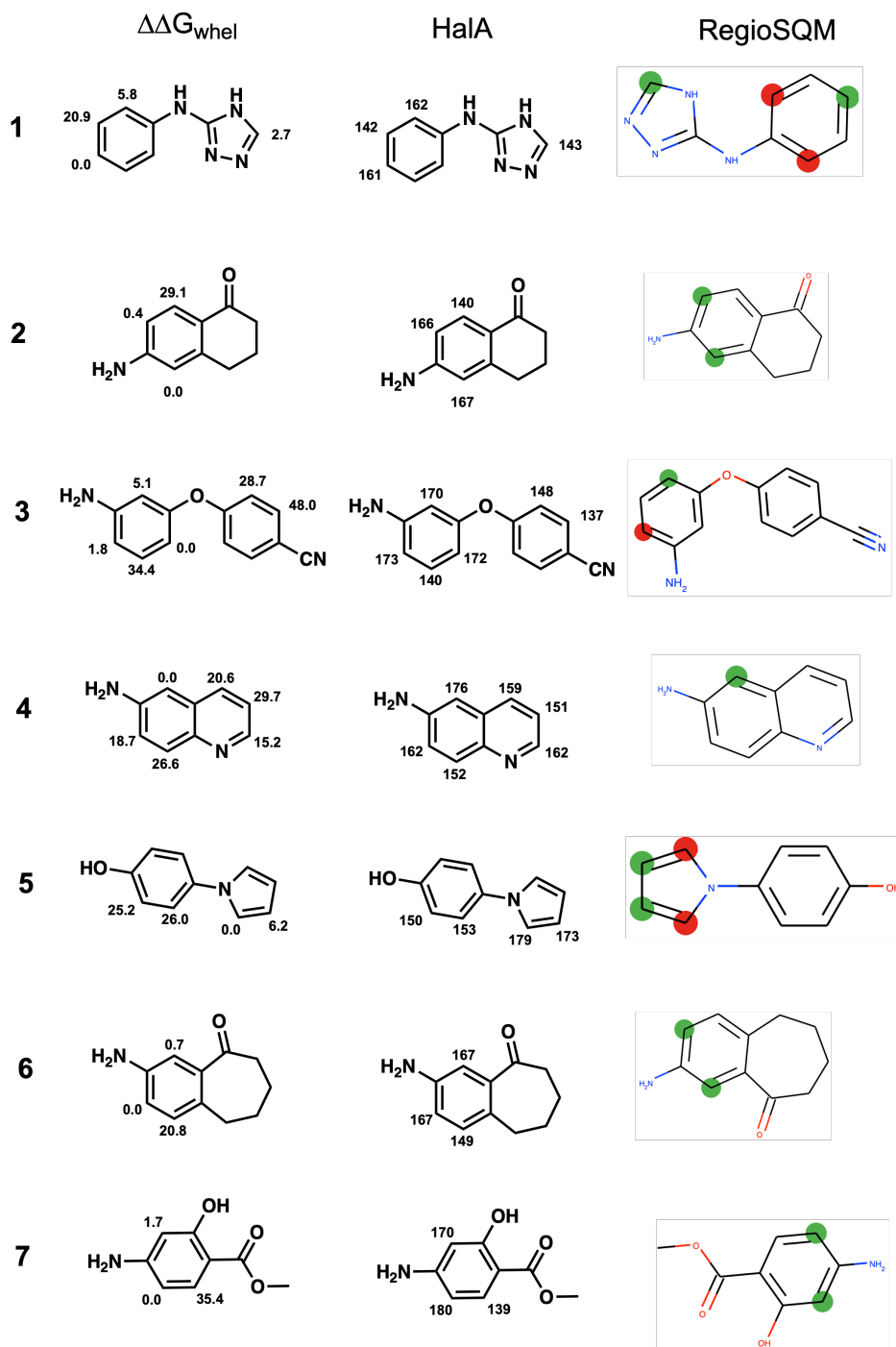
Comparison of the hydrogen bond triad in RebH (top) and OS (bottom) in simulations of the K79-336 HOCl complex. The bond between H109 and N470 is more persistent than the bond between H109 and S470 (right) but the H109-E357 interaction is unaffected (left).

Figure S3.15: Dynamics of substrate pi stacking



Comparison of pi-stacking residues H109 and F111 in 10S (blue) and 8F (green). While H109 is relatively unaffected, F111 has lower RMSD in 8F, indicating rigidification by the neighboring P110.

Figure S3.16: Comparison of methods



Comparison between $\Delta\Delta G_{\text{whel}}$ values used in this study with previously reported metrics HalA and RegioSQM. All three methods agree well when comparing the relative reactivity of each site.

3.8 Computational Models

Table S4.1: Computed Energies of Computational Models (Hartree)

Structure	Electronic Energy	Zero Point Energy	Enthalpy	Free Energy
Reactant Complex	-996.17861	0.221778	-995.94114	-995.997464
NAC5	-996.17076	0.221113	-995.93611	-995.9937
TS5	-996.14381	0.216471	-995.91225	-995.967078
NAC6	-996.17077	0.22101	-995.93611	-995.99385
TS6	-996.143661	0.216713	-995.91189	-995.968527
NAC7	-996.172	0.221115	-995.93751	-995.99517
TS7	-996.142862	0.21693	-995.91101	-995.965437

Cartesian Coordinates

Reactant Complex

C -1.93209 -0.43153 -1.44280
 C -1.03318 -0.88727 0.77806
 C -1.00364 -1.10555 -0.62935
 C -1.95257 -0.02373 1.38460
 C -2.84831 0.42992 -0.84506
 C -2.85610 0.63340 0.55534
 Cl 3.10263 0.20399 0.20867
 H -1.94893 0.13267 2.45905
 H -3.58030 1.31617 0.98982
 H 0.15044 3.07620 -1.21751
 H -3.57077 0.95736 -1.46125
 C 0.08596 -2.01457 -0.88397
 C 0.64764 -2.31526 0.34172
 N -0.02371 -1.64625 1.33466
 H 0.22545 -1.65888 2.31508
 H -1.92735 -0.57801 -2.51921
 H 0.36818 -2.44486 -1.83600
 H 1.47497 -2.96728 0.58380
 H 1.49747 -0.58966 -1.04492
 N 0.07899 2.26198 -0.59908
 H -0.86279 1.86223 -0.70786
 H 0.76098 1.54572 -0.91457
 C 0.33659 2.63909 0.82447
 H 0.30696 1.73368 1.42846

H 1.31811 3.10823 0.88583
 H -0.43775 3.33440 1.14688
 O 2.05349 0.23087 -1.16277

NAC5

C -1.12271200 -1.33602900 1.20752200
 C -1.80182400 0.18780300 -0.56920100
 C -1.50457100 -0.04201000 0.80603700
 C -1.73622500 -0.82750000 -1.53364100
 C -1.05078600 -2.34912000 0.24921700
 C -1.35902000 -2.09441000 -1.10850000
 Cl 1.73558700 -1.33608400 0.29320700
 H -1.97002900 -0.62361600 -2.57464300
 H -1.29493100 -2.90480900 -1.82879200
 H 3.78373900 -0.74112100 0.86782300
 H -0.77058700 -3.35396000 0.55163700
 C -1.66489300 1.21419000 1.48186200
 C -2.04013700 2.13355600 0.53226100
 N -2.13088500 1.51797000 -0.69993700
 H -2.37656500 1.97826200 -1.56577800
 H -0.89211400 -1.54411200 2.24857500
 H -1.51663500 1.40927200 2.53538300
 H -2.25830600 3.18797400 0.62916100
 O 3.36623100 -0.75282500 -0.01805000
 H 3.08457100 1.03013200 -0.41906000

N 2.74619800 1.99899900 -0.58872700
H 2.84287800 2.20045900 -1.58957300
H 3.36256900 2.64402000 -0.08340900
C 1.32530200 2.15644100 -0.14380000
H 1.01472700 3.18469200 -0.32485300
H 0.70783700 1.46608500 -0.71477900
H 1.27004000 1.92643000 0.91828800

TS5

C -0.68342000 -0.98231900 1.37088700
C -1.92068600 -0.23485800 -0.57388900
C -1.53929000 -0.07216700 0.80963300
C -1.48010600 -1.31381900 -1.38416900
C -0.09930800 -2.02134300 0.53681100
C -0.60823000 -2.19887700 -0.82340500
Cl 1.74856700 -1.13621100 0.23280500
H -1.82950700 -1.41228000 -2.40632800
H -0.22294400 -3.03521300 -1.39653100
H 3.97625700 0.02459800 0.78017200
H 0.20828800 -2.92876300 1.05167400
C -2.18362900 1.13445500 1.27940300
C -2.88477400 1.63552900 0.22755400
N -2.72152600 0.79626000 -0.88419800
H -3.14904600 0.94871700 -1.79151200
H -0.36504800 -0.90397500 2.40525600
H -2.11019400 1.55833300 2.27051600
H -3.49971600 2.51937300 0.14249600
O 3.69226700 -0.02865300 -0.14968400
H 3.07319100 1.14870500 -0.33322400
N 2.52434500 2.19524600 -0.52280700
H 2.60667600 2.42797500 -1.51531400
H 3.02908600 2.92282500 -0.01131400
C 1.09304000 2.18724300 -0.11845400
H 0.64821500 3.17414300 -0.26582900
H 0.56268700 1.44995700 -0.72031700
H 1.02696800 1.90681800 0.93286100

NAC6

C -1.68589200 -0.69970400 1.62527500
C -1.53323300 -0.07144000 -0.72755100
C -1.77384300 0.29312600 0.62898400
C -1.20851600 -1.38003100 -1.10523600
C -1.36438800 -1.99990800 1.25889900
C -1.12461800 -2.33959400 -0.09288700
Cl 1.68669100 -1.32855200 -0.39760700

H -1.03423800 -1.63795500 -2.14544200
H -0.88376400 -3.36622100 -0.35243700
H 3.65625500 -1.20820200 0.58688900
H -1.29121500 -2.77254200 2.01901300
C -2.05299300 1.70117200 0.63662500
C -1.97428400 2.13136200 -0.66615400
N -1.67082900 1.06711400 -1.48993000
H -1.53528900 1.12281400 -2.49034600
H -1.86587200 -0.44619300 2.66676600
H -2.28120300 2.31429800 1.49794300
H -2.11549900 3.11812000 -1.08477000
O 3.34747400 -0.77452300 -0.23536900
H 3.10074200 0.99232600 0.23144000
N 2.80295200 1.96471600 0.45035000
H 3.22396500 2.23173300 1.34606100
H 3.19276500 2.58231300 -0.26917400
C 1.31033200 2.08192200 0.50331800
H 0.93770000 1.40704300 1.27130000
H 1.05371500 3.11274500 0.74412600
H 0.90816000 1.80706500 -0.46966900

TS6

C 1.22772300 -0.88493500 -1.68528700
C 1.62598200 -0.32883300 0.69843100
C 1.85326400 -0.05820100 -0.70177400
C 0.80632600 -1.32386100 1.13751000
C 0.38399700 -1.87108900 -1.27350200
C 0.05553200 -2.06912700 0.13983800
Cl -1.76952200 -1.09859100 0.34068200
H 0.64159500 -1.51966300 2.19170100
H -0.23918900 -3.07610400 0.42562900
H -4.04245100 -0.17381400 -0.41358700
H -0.11971800 -2.50730500 -1.99357900
C 2.70495000 1.04989100 -0.77204800
C 2.99178000 1.42590400 0.54667500
N 2.36239400 0.60802900 1.41498500
H 2.39225500 0.68418500 2.42520300
H 1.43090700 -0.71723700 -2.73793700
H 3.08515900 1.53396600 -1.66020000
H 3.61445500 2.23419100 0.90550000
O -3.68092400 0.12799200 0.43836600
H -3.00963500 1.27331000 0.06946100
N -2.45397600 2.26480500 -0.23293500
H -2.92197900 2.64954600 -1.05715300
H -2.58412300 2.94553200 0.51938400

C -1.00307900 2.08140900 -0.50996700
H -0.88625100 1.35718900 -1.31584200
H -0.55087600 3.03290800 -0.79845400
H -0.51865100 1.69862400 0.38775500

NAC7

C -1.78839300 0.33895000 1.57466000
C -1.14315100 0.88557300 -0.71309600
C -1.03042700 1.12558800 0.68595800
C -1.96469400 -0.11973900 -1.23742100
C -2.61034500 -0.65571300 1.06112100
C -2.69246200 -0.89012400 -0.33067500
Cl 0.25946300 -2.10646600 -0.41872900
H -2.02636700 -0.29225100 -2.30780200
H -3.34060500 -1.67839200 -0.70270300
H 2.38157100 -0.80214600 0.70412000
H -3.19625700 -1.27289600 1.73649000
C -0.06816800 2.18335700 0.84771100
C 0.35142700 2.53849700 -0.41564500
N -0.29777900 1.76638300 -1.35258600
H -0.13846400 1.80472800 -2.35048400
H -1.72222000 0.50653200 2.64647400
H 0.24949700 2.64007900 1.77571000
H 1.05971900 3.29278400 -0.72794100
H 1.75908500 -3.02844700 0.89530800
N 2.56274800 0.20606300 0.86366900
H 1.65790000 0.69530100 0.96175900
H 3.07151000 0.30582900 1.74788500
C 3.33853500 0.78570500 -0.27415000
H 3.51480100 1.84213600 -0.07415000
H 4.28344100 0.24938600 -0.35667900
H 2.75169700 0.66555300 -1.18409200
O 1.88438900 -2.50974400 0.07326800

TS7

C -2.52850800 -0.16292800 1.17111600
C -1.09677900 0.26617100 -0.73298600
C -1.84484200 0.75002300 0.38047400
C -0.92184400 -1.14795400 -1.00588500
C -2.47774100 -1.55435600 0.87895600
C -1.74107900 -2.02905300 -0.17581600
Cl 0.97079200 -1.50248700 -0.32566600
H -0.80676300 -1.43920300 -2.04810300
H 3.13829100 -0.14386100 0.16882400
C -1.68469400 2.18383900 0.39915100

C -0.87710900 2.49830700 -0.65335900
N -0.53769900 1.32463700 -1.33678900
H 0.09942800 1.27699100 -2.12530500
H -3.10726600 0.18140500 2.02359400
H -2.12280100 2.87310900 1.10673000
H -0.50961000 3.44967100 -1.00845200
H 2.98909200 -1.58771200 1.28866800
N 2.96369800 1.00944600 0.00725100
H 3.85126700 1.50639600 -0.09143700
H 2.45075500 1.12058500 -0.87056000
C 2.17032900 1.56691400 1.13343200
H 1.95101500 2.62298200 0.96501100
H 1.23828300 1.00611000 1.21353500
H 2.74450600 1.45051700 2.05417200
O 3.16909000 -1.50083700 0.33569400
H -1.70475200 -3.09073000 -0.39473600
H -3.03601900 -2.24163800 1.50511000

MD starting position for K79-HOCl in RebH (2OA1) Chain A

C -25.46500 15.41800 0.28100
C -25.36000 16.90300 0.70700
C -26.32200 17.78500 -0.02500
C -25.69900 19.05500 -0.45900
C -25.99800 20.11600 0.56200
N -26.90400 21.13300 0.01500
H -26.74200 21.23000 -0.97700
H -26.73300 22.01800 0.47200
H -27.86100 20.85200 0.17300
H -26.54800 15.42600 0.40800
H -25.56600 16.97400 1.77500
H -24.34500 17.25400 0.51800
H -27.16000 18.01300 0.63400
H -26.69400 17.25400 -0.90200
H -26.10700 19.35100 -1.42500
H -24.62000 18.92300 -0.54300
H -26.46600 19.65300 1.43100
H -25.06600 20.59200 0.86800
O -26.13800 21.34000 -2.31100
H -25.79900 20.61700 -2.84400
Cl -24.88300 22.10700 -1.56200

MD starting position for K79-KCl in RebH (2OA1) Chain A

C -25.46500 15.41800 0.28100

C -25.36000 16.90300 0.70700
C -26.13400 17.81900 -0.18900
C -25.82300 19.24300 0.06300
C -26.32700 20.06100 -1.09300
N -25.89700 21.46000 -0.97500
H -26.50600 15.09600 0.28600
H -25.74400 17.00200 1.72200
H -24.31100 17.20100 0.69600
H -25.89000 17.58500 -1.22500
H -27.20000 17.65500 -0.02900
H -26.31200 19.56700 0.98200
H -24.74500 19.37000 0.15800
H -25.93500 19.64500 -2.02100
H -27.41600 20.02100 -1.11300
Cl -25.69500 22.11900 -2.51800
H -25.02100 21.50200 -0.47300
H -26.59800 21.98700 -0.47400

C -25.17300 105.30600 -16.82300
C -24.20300 105.96900 -15.89000
C -24.00500 107.40500 -16.23600
C -23.26400 108.09100 -15.12900
N -23.27300 109.54300 -15.33100
H -24.68700 103.27900 -16.34400
H -24.71100 105.21200 -17.80600
H -26.06700 105.92400 -16.90700
H -23.24400 105.45500 -15.95200
H -24.58200 105.89600 -14.87000
H -24.97500 107.88200 -16.37300
H -23.43000 107.48000 -17.15900
H -22.23300 107.73700 -15.11400
H -23.74000 107.85600 -14.17700
Cl -23.40500 110.31900 -13.83600
H -22.41500 109.82600 -15.78200
H -24.05900 109.79700 -15.91200

MD starting position for K79-HOCl in 0S (7JU0) Chain A

C -25.56800 103.92000 -16.32700
C -25.17300 105.30600 -16.82300
C -23.83000 105.71700 -16.29700
C -23.84600 107.11800 -15.78900
C -24.23700 108.04900 -16.89700
N -23.82200 109.42000 -16.58300
H -24.68700 103.27900 -16.34400
H -26.06800 105.92300 -16.90000
H -24.71600 105.21600 -17.80800
H -23.10800 105.63900 -17.08300
H -23.55100 105.06000 -15.50000
H -22.87200 107.37900 -15.43700
H -24.55200 107.19900 -14.98900
H -25.29900 108.02100 -17.02300
H -23.76500 107.73600 -17.80500
H -23.00900 109.66000 -17.13100
H -23.59700 109.48700 -15.60000
H -24.57500 110.05700 -16.80000
O -23.51000 109.69100 -13.86400
H -24.20300 110.22300 -13.46600
Cl -23.44600 108.25700 -12.90400

MD starting position for K79-HOCl in 10S (6PV2) Chain A

C 25.61700 -104.19400 -13.77100
C 25.24000 -105.60500 -14.23300
C 23.83700 -106.03900 -13.78900
C 23.85500 -107.44200 -13.22600
C 24.38000 -108.42700 -14.22500
N 23.82100 -109.77300 -13.90700
H 22.96900 -109.91600 -14.43000
H 24.49300 -110.48400 -14.15600
H 23.62200 -109.83000 -12.91800
H 26.64000 -104.52400 -13.59300
H 25.28200 -105.63400 -15.32200
H 25.96800 -106.31100 -13.83500
H 23.47800 -105.35200 -13.02200
H 23.16300 -106.00500 -14.64500
H 24.49200 -107.46100 -12.34200
H 22.84200 -107.72600 -12.94200
H 24.07400 -108.13000 -15.22800
H 25.46800 -108.46000 -14.16900
O 23.26900 -109.93100 -11.16500
H 22.36800 -110.23800 -11.03700
Cl 23.44600 -108.44100 -10.48200

MD starting position for K79-Cl in 0S (7JU0) Chain A

C -25.56800 103.92000 -16.32700

MD starting position for K79-HOCl in 8F (6P00) Chain A

C -20.03500 25.86700 -13.98200

C -21.44000 25.46900 -14.43200
C -22.48600 25.65100 -13.36600
C -23.61200 24.65300 -13.53300
C -24.55900 25.13800 -14.64400
N -25.77800 25.77300 -14.07800
H -25.69400 26.77800 -14.13800
H -25.88100 25.50100 -13.11100
H -26.58800 25.47100 -14.60100
H -20.25900 26.91600 -14.17700
H -21.71300 26.08100 -15.29100
H -21.42800 24.42200 -14.73700
H -22.02600 25.50900 -12.38800
H -22.89000 26.66100 -13.42900
H -23.20000 23.68100 -13.80400
H -24.16300 24.56600 -12.59700
H -24.03500 25.86600 -15.26400
H -24.85200 24.28800 -15.26100
O -26.06300 25.02000 -11.39700
H -26.81000 25.48500 -11.01100
Cl -24.68900 25.40300 -10.56800

**Tryptamine docked into active site of 10S
(6PV2) Chain A**

C 26.07500 -113.48100 -6.98300
C 26.48900 -112.74100 -5.90000
C 25.06000 -112.70400 -7.64800
C 24.91100 -111.49600 -6.90800
C 24.27000 -112.90100 -8.79500
N 25.79400 -111.55300 -5.85100
C 24.00300 -110.49800 -7.28200
C 23.36700 -111.91300 -9.16900
C 23.23400 -110.72500 -8.41900
H 25.93400 -110.82500 -5.16400
C 26.55100 -114.84700 -7.38000
C 27.98500 -114.78700 -7.90800
N 28.88300 -115.71800 -7.12400
H 29.38000 -116.36800 -7.74500
H 29.59500 -115.19600 -6.59800
H 28.35100 -116.28000 -6.44900

**Tryptamine docked into active site of 8F
(6P00) Chain A**

C -27.47400 21.06300 -6.69900
C -26.48800 21.49300 -5.84100
C -27.63800 22.09700 -7.68900

C -26.71200 23.13000 -7.36700
C -28.46500 22.25100 -8.81600
N -26.03100 22.73000 -6.23800
C -26.59800 24.29600 -8.13300
C -28.35500 23.40800 -9.57900
C -27.43200 24.42000 -9.23900
H -25.28900 23.24700 -5.78700
C -28.25800 19.78800 -6.60700
C -27.34700 18.57600 -6.81100
N -26.58100 18.69500 -8.11000
H -26.96100 19.44500 -8.70000
H -25.59100 18.91200 -7.94300
H -26.61800 17.82400 -8.65400

**Substrate 1 docked into active site of RebH
(2OA1) Chain A**

N -22.90600 24.49300 -5.97300
C -21.65200 24.30600 -6.49200
N -20.83100 25.37500 -6.73900
C -19.68200 24.83000 -7.24300
H -21.03400 26.35600 -6.58100
N -21.04500 23.17800 -6.82300
N -19.77800 23.52900 -7.30500
H -23.16100 25.45800 -5.79600
C -24.17500 22.42700 -6.46100
C -25.16700 21.52700 -6.06200
C -25.87400 21.70900 -4.87300
C -25.57900 22.81800 -4.07300
C -24.59500 23.72500 -4.45500
C -23.88200 23.53800 -5.65500

**Substrate 2 docked into active site of RebH
(2OA1) Chain A**

C -24.36500 23.48800 -5.05300
C -25.01000 22.36500 -5.58600
C -24.44200 21.67900 -6.66700
C -23.22900 22.11600 -7.21700
C -22.58100 23.24000 -6.68700
C -23.15200 23.92800 -5.60000
N -26.16700 21.94600 -5.06200
H -26.62700 21.14300 -5.44300
H -26.57300 22.43500 -4.29000
C -21.36500 23.66600 -7.25200
C -22.45300 25.16500 -5.00500
C -20.93600 25.02100 -5.18900

C -20.63100 24.89700 -6.68900
O -20.85600 23.01400 -8.25700
H -24.80300 24.01600 -4.21900
H -24.93900 20.81200 -7.07700
H -22.79300 21.58600 -8.05100
H -22.68200 25.23900 -3.96200
H -20.43300 25.89900 -4.78400
H -20.58800 24.12800 -4.67000
H -19.55700 24.78000 -6.83500
H -20.97300 25.79400 -7.20600
H -22.79300 26.04500 -5.50900

**Substrate 3 docked into active site of RebH
(2OA1) Chain A**

C -25.90300 22.56100 -4.07400
C -25.30300 22.23000 -5.30500
C -24.21400 23.00300 -5.75300
C -23.76900 24.06900 -4.98300
C -24.35700 24.41300 -3.76600
C -25.43000 23.63500 -3.32300
N -25.73000 21.12400 -6.03500
H -26.53800 21.29800 -6.62400
H -25.85400 20.28300 -5.48100
O -22.73800 24.85500 -5.51800
C -21.95800 24.35500 -6.53200
C -20.94000 25.08700 -7.15800
C -22.20900 23.02300 -6.89000
C -20.18100 24.48300 -8.15300
C -21.44900 22.42100 -7.88100
C -20.43000 23.15100 -8.52400
C -19.65000 22.53800 -9.55000
N -19.01100 22.03800 -10.38800

**Substrate 4 docked into active site of RebH
(2OA1) Chain A**

C -23.39500 23.11000 -6.19000
C -24.08700 23.79300 -5.18000
C -23.53300 24.94700 -4.61000
C -22.28800 25.41900 -5.04900
C -21.59600 24.73700 -6.05900
C -22.15200 23.58400 -6.62800
N -25.27500 23.34300 -4.76200
H -25.76800 23.83000 -4.04100
H -25.67000 22.52000 -5.16900
C -21.45600 22.91100 -7.63400

C -20.21900 23.41100 -8.04500
C -19.71000 24.56900 -7.44200
N -20.40300 25.19500 -6.47500
H -23.82000 22.22000 -6.63000
H -24.06600 25.47300 -3.83200
H -21.86200 26.30800 -4.60900
H -21.86700 22.02000 -8.08600
H -19.65900 22.91000 -8.82100
H -18.75400 24.96100 -7.75600

**Substrate 5 docked into active site of RebH
(2OA1) Chain A**

C -20.58100 24.71300 -5.40300
C -20.28200 24.64500 -6.77100
C -21.21700 24.11300 -7.66700
C -22.45300 23.64800 -7.19600
C -22.75200 23.71500 -5.82900
C -21.81700 24.24700 -4.93200
N -23.93200 23.27100 -5.38000
C -24.61600 23.79400 -4.34400
C -24.61900 22.23500 -5.89300
C -25.79600 23.06300 -4.17900
C -25.79900 22.07600 -5.16000
O -19.02100 25.11900 -7.25000
H -18.44700 24.37200 -7.43400

**Substrate 6 docked into active site of RebH
(2OA1) Chain A**

C -24.50500 24.23400 -4.49000
C -24.88100 23.04600 -5.14100
C -24.07400 22.53000 -6.16400
C -22.90600 23.21300 -6.51200
C -22.70800 24.51300 -6.08300
C -23.41200 24.97500 -4.96600
N -25.99700 22.40400 -4.77700
H -26.26000 21.55800 -5.23900
H -26.56900 22.77300 -4.04500
C -21.77700 25.49800 -6.85200
C -21.83100 22.53700 -7.33800
C -20.27400 25.29300 -6.62900
C -20.34500 22.72400 -6.92700
C -19.71900 24.04300 -7.38800
O -22.13500 21.79600 -8.30900

Substrate 7 docked into active site of RebH

(20A1) Chain A

C -24.49200 21.95600 -6.28800
C -25.29100 22.33400 -5.21000
C -25.01800 23.50000 -4.48000
C -23.90700 24.27200 -4.85300
C -23.10400 23.90000 -5.93100
C -23.40200 22.74400 -6.64500
N -26.41900 21.46600 -4.83800
H -27.05300 21.75200 -4.12100

H -26.54900 20.59000 -5.30200
O -22.57900 22.35900 -7.75100
H -21.95900 21.68400 -7.46900
C -21.89500 24.76700 -6.32800
O -21.65500 25.35800 -7.38000
O -20.98000 24.86400 -5.33100
C -20.14500 23.70300 -5.30000

References

- [1] Karl-Heinz van Pée and Eugenio P. Patallo. Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Applied Microbiology and Biotechnology*, 70(6):631–641, May 2006.
- [2] Changjiang Dong, Silvana Flecks, Susanne Unversucht, Caroline Haupt, Karl-Heinz van Pée, and James H. Naismith. Tryptophan 7-halogenase (prna) structure suggests a mechanism for regioselective chlorination. *Science*, 309(5744):2216–2219, 2005.
- [3] Ellen Yeh, Leah C. Blasiak, Alexander Koglin, Catherine L. Drennan, and Christopher T. Walsh. Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry*, 46(5):1284–1292, 2007. PMID: 17260957.
- [4] Silvana Flecks, Eugenio P. Patallo, Xiaofeng Zhu, Aliz J. Ernyei, Gotthard Seifert, Alexander Schneider, Changjiang Dong, James H. Naismith, and Karl-Heinz van Pée. New insights into the mechanism of enzymatic chlorination of tryptophan. *Angewandte Chemie International Edition*, 47(49):9533–9536, 2008.
- [5] Ajda Podgoršek, Marko Zupan, and Jernej Iskra. Oxidative halogenation with “green” oxidants: Oxygen and hydrogen peroxide. *Angewandte Chemie International Edition*, 48(45):8424–8450, 2009.
- [6] Jonathan Latham, Eileen Brandenburger, Sarah A. Shepherd, Binuraj R. K. Menon, and Jason Micklefield. Development of halogenase enzymes for use in synthesis. *Chemical Reviews*, 118(1):232–269, 2018. PMID: 28466644.
- [7] Veit Weichold, Daniela Milbredt, and Karl-Heinz van Pée. Specific enzymatic halogenation—from the discovery of halogenated enzymes to their applications in vitro and in vivo. *Angewandte Chemie International Edition*, 55(22):6374–6389, 2016.
- [8] Mary C. Andorfer, Jonathan E. Grob, Christine E. Hajdin, Julia R. Chael, Piro Siuti, Jeremiah Lilly, Kian L. Tan, and Jared C. Lewis. Understanding flavin-dependent halogenase reactivity via substrate activity profiling. *ACS Catalysis*, 7(3):1897–1904, 2017. PMID: 28989809.
- [9] Brian F. Fisher, Harrison M. Snodgrass, Krysten A. Jones, Mary C. Andorfer, and Jared C. Lewis. Site-selective c–h halogenation using flavin-dependent halogenases identified via family-wide activity profiling. *ACS Central Science*, 5(11):1844–1856, 2019. PMID: 31807686.
- [10] Mary C. Andorfer and Jared C. Lewis. Understanding and improving the activity of flavin-dependent halogenases via random and targeted mutagenesis. *Annual Review of Biochemistry*, 87(1):159–185, 2018. PMID: 29589959.
- [11] Mary C. Andorfer, Hyun June Park, Jaylie Vergara-Coll, and Jared C. Lewis. Directed evolution of rebh for catalyst-controlled halogenation of indole c–h bonds. *Chem. Sci.*, 7:3720–3729, 2016.

- [12] Tatyana G. Karabencheva-Christova, Juan Torras, Adrian J. Mulholland, Alessio Lodola, and Christo Z. Christov. Mechanistic insights into the reaction of chlorination of tryptophan catalyzed by tryptophan 7-halogenase. *Scientific Reports*, 7(1):17395, Dec 2017.
- [13] Jon Ainsley, Adrian J. Mulholland, Gary W. Black, Olivier Sparagano, Christo Z. Christov, and Tatyana G. Karabencheva-Christova. Structural insights from molecular dynamics simulations of tryptophan 7-halogenase and tryptophan 5-halogenase. *ACS Omega*, 3(5):4847–4859, 2018. PMID: 31458701.
- [14] Eduard Bitto, Yu Huang, Craig A. Bingman, Shanteri Singh, Jon S. Thorson, and George N. Phillips Jr. The structure of flavin-dependent tryptophan 7-halogenase rebh. *Proteins: Structure, Function, and Bioinformatics*, 70(1):289–293, 2008.
- [15] Lizbeth Hedstrom. Serine protease mechanism and specificity. *Chemical Reviews*, 102(12):4501–4524, 2002. PMID: 12475199.
- [16] Magnus Liljenberg, Tore Brinck, Björn Herschend, Tobias Rein, Glen Rockwell, and Mats Svensson. Validation of a computational model for predicting the site for electrophilic substitution in aromatic systems. *The Journal of Organic Chemistry*, 75(14):4696–4705, 2010. PMID: 20552984.
- [17] Xiaofeng Zhu, Walter De Laurentis, Khim Leang, Julia Herrmann, Katja Ihlefeld, Karl-Heinz van Pée, and James H. Naismith. Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *Journal of Molecular Biology*, 391(1):74–85, 2009.
- [18] Ann-Christin Moritzer, Hannah Minges, Tina Prior, Marcel Frese, Norbert Sewald, and Hartmut H. Niemann. Structure-based switch of regioselectivity in the flavin-dependent tryptophan 6-halogenase thal. *Journal of Biological Chemistry*, 294(7):2529–2542, 2019.
- [19] Kazi Lingkon and John J. Bellizzi III. Structure and activity of the thermophilic tryptophan-6 halogenase borh. *ChemBioChem*, 21(8):1121–1128, 2020.
- [20] Thomas C. Bruice and Felice C. Lightstone. Ground state and transition state contributions to the rates of intramolecular and enzymatic reactions. *Accounts of Chemical Research*, 32(2):127–136, 1999.
- [21] Hsiao-Lin Lee, Chih-Kang Chang, Wen-Yih Jeng, Andrew H.-J. Wang, and Po-Huang Liang. Mutations in the substrate entrance region of β -glucosidase from *Trichoderma reesei* improve enzyme activity and thermostability. *Protein Engineering, Design and Selection*, 25(11):733–740, 10 2012.
- [22] Susanne Zehner, Alexander Kotsch, Bojan Bister, Roderich D. Süssmuth, Carmen Méndez, José A. Salas, and Karl-Heinz van Pée. A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* ll-42d005. *Chemistry & Biology*, 12(4):445–452, 2005.
- [23] Fang-Yuan Chang and Sean F. Brady. Cloning and characterization of an environmental dna-derived gene cluster that encodes the biosynthesis of the antitumor substance be-54017. *Journal of the American Chemical Society*, 133(26):9996–9999, 2011. PMID: 21542592.

- [24] Katherine S. Ryan. Biosynthetic gene cluster for the cladoniamides, bis-indoles with a rearranged scaffold. *PLOS ONE*, 6(8):1–10, 08 2011.
- [25] Weslee S. Glenn, Ezekiel Nims, and Sarah E. O'Connor. Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *Journal of the American Chemical Society*, 133(48):19346–19349, 2011. PMID: 22050348.
- [26] Sarah A. Shepherd, Binuraj R. K. Menon, Heidi Fisk, Anna-Winona Struck, Colin Levy, David Leys, and Jason Micklefield. A structure-guided switch in the regioselectivity of a tryptophan halogenase. *ChemBioChem*, 17(9):821–824, 2016.
- [27] Jia Zeng and Jixun Zhan. Characterization of a tryptophan 6-halogenase from streptomyces toxytricini. *Biotechnology Letters*, 33(8):1607–1613, Aug 2011.
- [28] John R. Heemstra and Christopher T. Walsh. Tandem action of the o2- and fadh2-dependent halogenases ktzq and ktzr produce 6,7-dichlorotryptophan for kutzneride assembly. *Journal of the American Chemical Society*, 130(43):14024–14025, 2008. PMID: 18828589.
- [29] Philip A. Romero and Frances H. Arnold. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology*, 10(12):866–876, Dec 2009.
- [30] Amy E. Fraley, Marc Garcia-Borras, Ashootosh Tripathi, Dheeraj Khare, Eduardo V. Mercado-Marin, Hong Tran, Qingyun Dan, Gabrielle P. Webb, Katharine R. Watts, Phillip Crews, Richmond Sarpong, Robert M. Williams, Janet L. Smith, K. N. Houk, and David H. Sherman. Function and structure of mala/mala', iterative halogenases for late-stage c–h functionalization of indole alkaloids. *Journal of the American Chemical Society*, 139(34):12060–12068, Aug 2017.
- [31] Aisaraphon Phintha, Kridsakorn Prakinee, Aritsara Jaruwat, Narin Lawan, Surawit Visitsatthawong, Chadaporn Kantiwiriyanitch, Warangkhan Songsunthong, Duangthip Trisrivirat, Pirom Chenprakhon, Adrian Mulholland, et al. Dissecting the low catalytic capability of flavin-dependent halogenases. *Journal of Biological Chemistry*, 296, 2021.
- [32] Jimmy C Kromann, Jan H Jensen, Monika Kruszyk, Mikkel Jessing, and Morten Jørgensen. Fast and accurate prediction of the regioselectivity of electrophilic aromatic substitution reactions. *Chemical science*, 9(3):660–665, 2018.
- [33] Kumar Dilip Ashtekar, Nastaran Salehi Marzijarani, Arvind Jaganathan, Daniel Holmes, James E Jackson, and Babak Borhan. A new tool to guide halofunctionalization reactions: the halenium affinity (hala) scale. *Journal of the American Chemical Society*, 136(38):13355–13362, 2014.
- [34] Dibyendu Mondal, Brian F Fisher, Yuhua Jiang, and Jared C Lewis. Flavin-dependent halogenases catalyze enantioselective olefin halocyclization. *Nature communications*, 12(1):1–7, 2021.
- [35] Ellen Yeh, Sylvie Garneau, and Christopher T Walsh. Robust in vitro activity of rebf and rebh, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proceedings of the National Academy of Sciences*, 102(11):3960–3965, 2005.

- [36] Lisa D Cabrita, Weiwen Dai, and Stephen P Bottomley. A family of e. coli expression vectors for laboratory scale and high throughput soluble protein production. *BMC biotechnology*, 6(1):1–8, 2006.
- [37] Joseph Sambrook. A laboratory manual. *Molecular cloning*, 1, 2001.
- [38] Karin L Heckman and Larry R Pease. Gene splicing and mutagenesis by pcr-driven overlap extension. *Nature protocols*, 2(4):924–932, 2007.
- [39] James T Payne, Mary C Andorfer, and Jared C Lewis. Regioselective arene halogenation using the fad-dependent halogenase rebh. *Angewandte Chemie International Edition*, 52(20):5271–5274, 2013.
- [40] Mary C Andorfer, Hyun June Park, Jaylie Vergara-Coll, and Jared C Lewis. Directed evolution of rebh for catalyst-controlled halogenation of indole c–h bonds. *Chemical science*, 7(6):3720–3729, 2016.
- [41] Ellen Yeh, Leah C Blasiak, Alexander Koglin, Catherine L Drennan, and Christopher T Walsh. Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry*, 46(5):1284–1292, 2007.
- [42] Catherine B Poor, Mary C Andorfer, and Jared C Lewis. Improving the stability and catalyst lifetime of the halogenase rebh by directed evolution. *ChemBioChem*, 15(9):1286–1289, 2014.
- [43] WM Zbyszek Otwinowski. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol*, 276:307–326, 1997.
- [44] AJ McCoy, LC Storoni, G Bunkoczi, RD Oeffner, RJ Read, RW Grosse-Kunstleve, PD Adams, and MD Winn. Phaser (ccp4: Supported program) name phaser-2.5. 0-maximum likelihood analysis and phasing. synopsis phaser. *J. Appl. Cryst*, 40:658–674, 2007.
- [45] Eduard Bitto, Yu Huang, Craig A Bingman, Shanteri Singh, Jon S Thorson, and George N Phillips Jr. The structure of avin-dependent tryptophan 7-halogenase rebh. *Proteins: Structure, Function, and Bioinformatics*, 70(1), 2008.
- [46] P Emsley, B Lohkamp, WG Scott, and K Cowtan. Features and development of coot. *biol. crystallogr*, 2010.
- [47] Paul D Adams, Ralf W Grosse-Kunstleve, L-W Hung, Thomas R Ioerger, Airlie J McCoy, Nigel W Moriarty, Randy J Read, James C Sacchettini, Nicholas K Sauter, and Thomas C Terwilliger. Phenix: building new software for automated crystallographic structure determination. *Acta Crystallographica Section D: Biological Crystallography*, 58(11):1948–1954, 2002.
- [48] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, G Scalmani, V Barone, GA Petersson, H Nakatsuji, et al. Gaussian 16 revision c. 01. 2016. 421, 2016.

- [49] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gauther, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D Peterson, et al. Xsede: accelerating scientific discovery. *Computing in science & engineering*, 16(5):62–74, 2014.
- [50] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993.
- [51] Chengteh Lee, Weitao Yang, and Robert G Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical review B*, 37(2):785, 1988.
- [52] V Myrseth, JD Bozek, E Kukk, LJ Sæthre, and TD Thomas. Adiabatic and vertical carbon 1s ionization energies in representative small molecules. *Journal of electron spectroscopy and related phenomena*, 122(1):57–63, 2002.
- [53] Philip J Stephens, Frank J Devlin, Cary F Chabalowski, and Michael J Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of physical chemistry*, 98(45):11623–11627, 1994.
- [54] Stephan Ehrlich, Jonas Moellmann, Werner Reckien, Thomas Bredow, and Stefan Grimme. System-dependent dispersion coefficients for the dft-d3 treatment of adsorption processes on ionic surfaces. *ChemPhysChem*, 12(17):3414–3420, 2011.
- [55] Aleksandr V Marenich, Christopher J Cramer, and Donald G Truhlar. 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. *The Journal of Physical Chemistry B*, 113(18):6378–6396, 2009.
- [56] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics*, 10(44):6615–6620, 2008.
- [57] Douglas BG Mateus, Ana Paula L Batista, Renata L Rodrigues, and Sofia Nikolaou. No photorelease from a ruthenium complex assisted by formation of a supramolecular dimer. *Journal of the Brazilian Chemical Society*, 31:2319–2330, 2020.
- [58] Tobias Risthaus and Stefan Grimme. Benchmarking of london dispersion-accounting density functional theory methods on very large molecular complexes. *Journal of chemical theory and computation*, 9(3):1580–1591, 2013.
- [59] Yi-Pei Li, Joseph Gomes, Shaama Mallikarjun Sharada, Alexis T Bell, and Martin Head-Gordon. Improved force-field parameters for qm/mm simulations of the energies of adsorption for molecules in zeolites and a free rotor correction to the rigid rotor harmonic oscillator model for adsorption enthalpies. *The Journal of Physical Chemistry C*, 119(4):1840–1850, 2015.
- [60] Hui Li, Andrew D. Robertson, and Jan H. Jensen. Very fast empirical prediction and rationalization of protein pka values. *Proteins: Structure, Function, and Bioinformatics*, 61(4):704–721, 2005.

- [61] Delphine C. Bas, David M. Rogers, and Jan H. Jensen. Very fast prediction and rationalization of pka values for protein–ligand complexes. *Proteins: Structure, Function, and Bioinformatics*, 73(3):765–783, 2008.
- [62] Mats H. M. Olsson, Chresten R. Søndergaard, Michał Rostkowski, and Jan H. Jensen. Propka3: Consistent treatment of internal and surface residues in empirical pka predictions. *Journal of Chemical Theory and Computation*, 7(2):525–537, 2011. PMID: 26596171.
- [63] Chresten R. Søndergaard, Mats H. M. Olsson, Michał Rostkowski, and Jan H. Jensen. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pka values. *Journal of Chemical Theory and Computation*, 7(7):2284–2295, 2011. PMID: 26606496.
- [64] Oleg Trott and Arthur 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(2):455–461, 2010.
- [65] Andreas W. Götz, Mark J. Williamson, Dong Xu, Duncan Poole, Scott Le Grand, and Ross C. Walker. Routine microsecond molecular dynamics simulations with amber on gpus. 1. generalized born. *Journal of Chemical Theory and Computation*, 8(5):1542–1555, 2012. PMID: 22582031.
- [66] David A Case, H Metin Aktulga, Kellon Belfon, Ido Ben-Shalom, Scott R Brozell, David S Cerutti, Thomas E Cheatham III, Vinícius Wilian D Cruzeiro, Tom A Darden, Robert E Duke, et al. Amber 2021. 2021.
- [67] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.
- [68] Christopher I. Bayly, Piotr Cieplak, Wendy Cornell, and Peter A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993.
- [69] Brent H. Besler, Kenneth M. Merz Jr., and Peter A. Kollman. Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry*, 11(4):431–439, 1990.
- [70] U. Chandra Singh and Peter A. Kollman. An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry*, 5(2):129–145, 1984.
- [71] Andrej Šali and Tom L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, 1993.
- [72] Garrett M. Morris, Ruth Huey, William Lindstrom, Michel F. Sanner, Richard K. Belew, David S. Goodsell, and Arthur J. Olson. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16):2785–2791, 2009.

- [73] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffrey D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.
- [74] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015. PMID: 26574453.
- [75] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.
- [76] Daniel R. Roe and Thomas E. III Cheatham. Ptraj and cpptraj: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, 2013. PMID: 26583988.
- [77] Schrodinger. The pymol molecular graphics system, version 2.3. 2010.

Chapter 4

De Novo Design of Luciferases Using Deep Learning

4.1 Author List and Affiliation

Hsien-Wei (Andy) Yeh^{1,2,†}, Christoffer Norn^{1,2,†}, Yakov Kipnis^{1,2,5}, Doug Tischer^{1,2}, Samuel J. Pellock^{1,2}, Declan Evans³, Pengchen Ma^{3,4}, Gyu Rie Lee^{1,2}, Jason Z. Zhang^{1,2}, Ivan Anishchenko^{1,2}, Brian Coventry^{1,2}, Longxing Cao^{1,2}, Justas Dauparas^{1,2}, Samer Halabiya², Michelle DeWitt², Lauren Carter², K. N. Houk³ & David Baker^{1,2,5,*}

¹Department of Biochemistry, University of Washington, Seattle, Washington, USA

²Institute for Protein Design, University of Washington, Seattle, Washington, USA

³Department of Chemistry and Biochemistry, University of California, Los Angeles, USA

⁴School of Chemistry, Xi'an Key Laboratory of Sustainable Energy Materials Chemistry, MOE Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, Xi'an Jiaotong University, Xi'an, China

⁵Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA

[†]These authors contributed equally

*Corresponding author

This work is a collaboration between the groups of David Baker (Washington) and K. N. Houk (UCLA) and has been accepted at Nature as of 03 January 2023 (Nature manuscript 2022-01-00233C). Protein design, expression, and characterization were led by Hsien-Wei Yeh under the guidance of David Baker. DFT calculations were performed by Pengchen Ma and MD simulations were performed by Declan Evans under the guidance of K. N. Houk.

4.2 Abstract

De novo enzyme design has sought to introduce active sites and substrate binding pockets predicted to catalyze a reaction of interest into geometrically compatible native scaffolds [1, 2], but has been limited by a lack of suitable protein structures and the complexity of native protein sequence-structure relationships. Here we describe a deep-learning based “family-wide halluci-

nation” approach that generates large numbers of idealized protein structures containing diverse pocket shapes and designed sequences that encode them. We use these scaffolds to design artificial luciferases that selectively catalyze the oxidative chemiluminescence of the synthetic luciferin substrates diphenylterazine (DTZ) [3] and 2-deoxycoelenterazine (h-CTZ) through the placement of an arginine guanidinium group adjacent to an anion species that develops during the reaction in a high shape complementarity binding pocket. For both luciferin substrates, we obtain designed luciferases with high selectivity; the most active of these is a small (13.9 kDa) and thermostable ($T_M > 95^\circ\text{C}$) enzyme with a catalytic efficiency on DTZ ($k_{\text{cat}}/K_M = 106 \text{ M}^{-1}\text{s}^{-1}$) comparable to native luciferases but with much higher substrate specificity. The design of highly active and specific biocatalysts from scratch with broad applications in biomedicine is an important milestone for computational enzyme design, and our approach should enable the design of a wide range of novel luciferases and other enzymes.

4.3 Main

Bioluminescent light produced by the enzymatic oxidation of a luciferin substrate by luciferases is widely used for bioassays and imaging in biomedical research. Because no excitation light source is needed, luminescent photons are produced in the dark which results in higher sensitivity than fluorescence imaging in live animal models and in biological samples where autofluorescence or phototoxicity is a concern [4, 5]. However, the development of luciferases as molecular probes has lagged behind that of well-developed fluorescent protein toolkits for a number of reasons: (i) very few native luciferases have been identified [6, 7]; (ii) many of those that have been identified require multiple disulfide bonds to stabilize the structure and are therefore prone to misfolding in mammalian cells [8]; (iii) most native luciferases do not recognize synthetic luciferins with more desirable photophysical properties [9]; and (iv) multiplexed imaging to follow multiple processes in parallel using mutually orthogonal luciferase-luciferin pairs has been limited by the low substrate specificity of native luciferases [10, 11].

We sought to use de novo protein design to create new luciferases that are small, highly stable, well-expressed in cells, specific for one substrate, and need no cofactors to function. We chose a synthetic luciferin, diphenylterazine (DTZ) as the target substrate because of its high quantum yield, red-shifted emission [3], favorable in vivo pharmacokinetics [12, 13], and lack of required cofactors for light emission. Previous computational enzyme design studies have primarily repurposed native protein scaffolds in the PDB [1, 2], but there are few native structures with binding pockets appropriate for DTZ, and the effects of sequence changes on native proteins can be unpredictable (designed helical bundles have also been used as enzyme scaffolds [14, 15, 16], but these are limited in number and most do not have pockets appropriate for DTZ binding). To circumvent these limitations, we set out to generate large numbers of ideal protein scaffolds with pockets of the appropriate size and shape for DTZ, and with clear sequence-structure relationships to facilitate subsequent active site incorporation. To identify protein folds capable of hosting such pockets, we first docked DTZ into 4000 native small molecule binding proteins. We found that many NTF2 (nuclear transport factor 2)-like folds have binding pockets with appropriate shape-complementary and size for DTZ placement (red labels in Figure 4.1e), and hence selected the NTF2-like superfamily as the target topology.

4.3.1 Family-wide hallucination

Native NTF2 structures have a range of pocket sizes and shapes but also contain non-ideal features such as long loops which compromise stability. To create large numbers of ideal NTF2-like structures, we developed a deep-learning based “family-wide hallucination” approach that integrates unconstrained de novo design [17, 18] and fixed backbone sequence design approaches [19] to enable the generation of an essentially unlimited number of proteins having a desired fold (Figure 4.1a). The family-wide hallucination approach utilizes the de novo sequence and structure discovery capability of unconstrained protein hallucination [17, 18] for loop and variable regions, and structure-guided sequence optimization for core regions. We employed the trRosetta structure prediction neural network [20], which is effective in identifying experimentally successful de novo

designed proteins and hallucinating new globular proteins of diverse topologies. Starting from sequences and predicted structures of 2,000 naturally occurring NTF2s, we used trRosetta to optimize the amino sequence of conserved core and variable loop regions. Protein core idealization was carried out with a topology-specific loss function over core residue pair geometries (see Methods) and variable loop optimization, by optimizing sequence length and identity to maximize the confidence of the neural network in the predicted structure. To further encode structural specificity, we incorporated buried, long-range hydrogen-bonding networks. The resulting 1615 family-wide hallucinated NTF2 scaffolds provided more shape complementary binding pockets for DTZ than native small-molecule protein binding proteins (Figure 4.1e). This approach samples protein backbones closer to native NTF2-like proteins (Figure 4.1e) and with better scaffold quality metrics than a previous non deep-learning approach [21] (Figure 4.1e).

4.3.2 De novo design of luciferases for DTZ

Computational enzyme design generally starts from an ideal active site or theozyme consisting of protein functional groups surrounding the reaction transition state that is then matched into a set of existing scaffolds [1, 2]. However, the detailed catalytic geometry of marine luciferases is not well defined as only a handful of apo-structures and no holo-structures with luciferin substrates have been solved [22, 23, 24] (excluding calcium-regulated photoproteins). Both quantum chemistry calculations [25, 26] and experimental data [27, 28] suggest that the chemiluminescent reaction proceeds through an anionic species and that the polarity of the surroundings can substantially alter the free energy of the subsequent single electron transfer (SET) process with triplet molecular oxygen ($^3\text{O}_2$). Guided by these data (Figure E4.1), we sought to design a shape complementary catalytic site that stabilizes the anionic state of DTZ and lowers the SET energy barrier, assuming that the downstream dioxetane light emitter thermolysis steps are spontaneous. To stabilize the anionic state, we focused on the placement of the positively charged guanidinium group of an arginine residue to stabilize the developing negative charge on the imidazopyrazinone group.

To computationally design such active sites into large numbers of hallucinated NTF2 scaffolds,

we first generated an ensemble of anionic DTZ conformers (Figure 4.1b). Next, around each conformer, we used the RIFgen method [29, 30] to enumerate Rotamer Interaction Fields (RIFs) on 3D grids consisting of millions of placements of amino acid sidechains making hydrogen bonding and nonpolar interactions with DTZ (Figure 4.1c). An arginine guanidinium group was placed adjacent to the N1 atom of the imidazopyrazinone group to stabilize the developing negative charge. RIFdock was then used to dock each DTZ conformer and associated RIF in the central cavity of each scaffold to maximize protein-DTZ interactions. An average of eight sidechain rotamers including an arginine to stabilize the anionic imidazopyrazinone core were positioned in each pocket (Figure S4.1a). For the top 50,000 docks with the most favorable sidechain-DTZ interactions, we optimized the remainder of the sequence using RosettaDesign (Figure 4.1d) for high-affinity binding to DTZ with a bias towards the naturally observed sequence variation to ensure foldability. During the design process, pre-defined hydrogen bond networks (HBNet) in the scaffolds were kept intact for structural specificity and stability, and interactions of these HBNet side chains with DTZ were explicitly required in the RIFdock step to ensure preorganization of residues essential for catalysis. In a first sequence design step, the identities of all RIF and HBNet residues were kept fixed, and the surrounding residues were optimized to hold the sidechain-DTZ interactions in place and maintain structural specificity. In a second sequence design step, the RIF residue identities (except the arginine) were also allowed to vary, as Rosetta can identify apolar and aromatic packing interactions missed in the RIF due to binning effects. During sequence design, the scaffold backbone, sidechains, and DTZ substrate were allowed to relax in cartesian space. Following sequence optimization, the designs were filtered based on ligand-binding energy, protein-ligand hydrogen bonds, shape complementarity, and contact molecular surface, and 7648 designs were selected and ordered as pooled oligos for experimental screening.

4.3.3 Identification of active luciferases

Oligonucleotides encoding the two halves of each design were assembled into full-length genes and cloned into an *E. coli* expression vector (see Methods). A colony-based screening method was

used to directly image active luciferase colonies from the library and the activities of selected clones were confirmed using a 96-well plate expression (Figure E4.2). Three active designs were identified; we refer to the most active of these as LuxSit (Latin: let light exist) which at 117 residues (13.9 kDa) is smaller than any previously described luciferase. Biochemical analysis, including SDS-PAGE and size exclusion chromatography (Figure 4.2ab and E4.3), indicated that LuxSit is highly expressed in E coli and soluble and monomeric. Circular dichroism (CD) spectroscopy showed a strong far UV CD signature, suggesting an organized $\alpha - \beta$ structure. CD melting experiments showed that the protein is not fully unfolded at 95°C, and the full structure is regained when the temperature is dropped (Figure 4.2c). Incubation of LuxSit with DTZ resulted in luminescence with an emission peak at roughly 480 nm (Figure 4.2d), consistent with the DTZ chemiluminescence spectrum. While we were not able to determine the crystal structure of LuxSit, the AlphaFold2 [31] predicted structure is very close to the design model at the backbone level (RMSD = 1.3 Å) and over the side chains interacting with the substrate (Figure 4.2e). The designed LuxSit active site contains Tyr14-His98 and Asp18-Arg65 dyads; with the imidazole nitrogen atoms of His98 making hydrogen bond interactions with Tyr14 and the O1 atom of DTZ (Figure 4.2f). The center of the Arg65 guanidinium cation is 4.2 Å from the N1 atom of DTZ and Asp18 forms a bidentate hydrogen bond to the guanidinium group and backbone N-H of Arg65 (Figure 4.2g).

4.3.4 De novo design of luciferases for h-CTZ

We next sought to apply the knowledge gained from designing LuxSit to design 2-deoxycoelenterazine (h-CTZ) specific luciferases. Since the molecular shape of h-CTZ is different from that of DTZ, we created an additional set of NTF2 superfamily scaffolds (see Methods) with matching pocket shapes and high model confidence (AlphaFold2 predicted local distance difference test, pLDDT > 92). We then installed catalytic sites in these scaffolds and designed the first shell protein sidechain-h-CTZ interactions using the histidine and arginine substrate interaction geometries that were most successful in the first round for DTZ. To design the remainder of the sequence, we used

ProteinMPNN [32], which can result in better stability, solubility, and accuracy than RosettaDesign. Following filtering based on the AlphaFold2 predicted pLDDT, C_{α} RMSD, contact molecular surface, and Rosetta computed binding energies (see Methods), we selected and experimentally expressed 46 designs in *E. coli* and identified two (HTZ3-D2 and HTZ3-G4) with luciferase activity with the h-CTZ luciferin substrate. Both designs were highly soluble, monodisperse, and monomeric; the luciferase activities were of the same order of magnitude as LuxSit (Figure E4.4). The success rate increased from 3/7648 to 2/46 sequences in the second round likely due to the more optimal active site geometry and the increased robustness of ProteinMPPN sequence design.

4.3.5 Luciferase activity optimization

To better understand the contributions to the catalysis of LuxSit, the most active of our designs, we constructed a site saturation mutagenesis (SSM) library in which each residue in the substrate binding pocket was mutated to every other amino acid one at a time (see Methods), and determined the effect of each mutation to luciferase activity. Figure 4.2f-i illustrate the amino-acid preferences at key positions. Arg65 is highly conserved, and its dyad partner Asp18 can only be mutated to Glu (which reduces activity), suggesting the carboxylate-Arg65 hydrogen bond is important for luciferase activity. In the Tyr14-His98 dyad, Tyr14 can be substituted with Asp and Glu, while His98 can be replaced with Asn. As all active variants had hydrogen bond donors and acceptors at these positions, the dyads may help mediate the electron and proton transfer required for luminescence. Hydrophobic (Figure 4.2h) and π -stacking (Figure 4.2i) residues at the binding interface tolerate other aromatic or aliphatic substitutions and generally prefer the amino acid in the original design consistent with model-based affinity predictions of mutational effects (Figure E4.5). The A96M and M110V mutants (highlighted in pink) increase activity by 16-fold and 19-fold over LuxSit respectively (Table ??). Optimization guided by these results yielded LuxSit-f (A96M/M110V) with strong initial flash emission and LuxSit-i (R60S/A96L/M110V) with more than 100-fold higher photon flux over LuxSit (Figure E4.6). Overall, the active site saturation mutagenesis results support the design model, with the Tyr14-His98 and Asp18-Arg65 dyads playing key roles in catalysis

and the substrate-binding pocket largely conserved.

The most active catalysts, LuxSit-i (Figure E4.3beh) and LuxSit-f (Figure E4.3cfi) were both expressed solubly in *E. coli* at high levels and are monomeric (some dimerization was observed at the high protein concentration, (Figure E4.3l) and thermostable (Figure E4.3j-k). Similar to native CTZ-utilizing luciferases, the apparent Michaelis constants K_M of both LuxSit-i and LuxSit-f are in the low μM range (Figure 4.3a) and the luminescent signal decays over time due to fast catalytic turnover (Figure E4.7a). LuxSit-i is a very efficient enzyme with a k_{cat}/K_M of $106 \text{ M}^{-1}\text{s}^{-1}$. The luminescence signal is readily visible to the naked eye (Figure 4.3b), and the photon flux (photon s^{-1}) is 38% greater than the native *Renilla reniformis* luciferase (RLuc) (Table S2). The DTZ luminescent reaction catalyzed by LuxSit-i is pH-dependent (Extended Data Fig. 7b), consistent with the proposed mechanism. We used a combination of density functional theory (DFT) calculations and molecular dynamics (MD) simulations to investigate the basis for LuxSit activity in more detail; the results support the anion stabilization mechanism (Figure E4.8a and S4.2a) and suggest that LuxSit-i provides better DTZ transition state charge stabilization than LuxSit (Figure E4.8b).

4.3.6 Cell imaging and multiplexed bioassay

As luciferases are commonly used genetic tags and reporters for cell biological studies, we evaluated the expression and function of LuxSit-i in live mammalian cells. LuxSit-i-mTagBFP2-expressing HEK293T cells showed DTZ-specific luminescence (Figure 4.3c), which was maintained following targeting of LuxSit-i to the nucleus, membrane, and mitochondria (Figure E4.9). Native and previously engineered luciferases are quite promiscuous with activity on many luciferin substrates (Figure 4.4ac and Figure S4.3), possibly due to their large and open pockets (a luciferase with high specificity to one luciferin substrate has been difficult to control even with extensive directed evolution [33, 34]). In contrast, LuxSit-i exhibited exquisite specificity to its target luciferin with 50-fold selectivity for DTZ over bis-CTZ (which differs only in one benzylic carbon; MD simulations suggest this arises from greater transition state shape complementarity (Figure E4.8bc and S4.2bc)), 28-fold selectivity over 8pyDTZ (differing only in one nitrogen atom), and more

than 100-fold selectivity over other luciferin substrates (Figure 4.4b). Our most active design for h-CTZ (HTZ3-G4) was also highly specific for its target substrate (Figure 4.4c and E4.4d). Overall, the specificity of our designed luciferases is much greater than native luciferases [35, 36] or previously engineered luciferases [37] (Table ??).

The high substrate specificity of LuxSit-i might allow the multiplexing of luminescent reporters through substrate-specific or spectrally resolved luminescent signals (Figure 4.4d and E4.10ab). To explore this possibility, we tracked two independent signaling pathways (cAMP/PKA and NF- κ B) by placing the expression of either RLuc or LuxSit-i downstream of the NF- κ B or cAMP response element promoters, respectively (Figure 4.4e). Imaging in the presence of the substrates for the two luciferases (PP-CTZ for RLuc and DTZ for LuxSit-i) one at a time (Figure 4.4f) can clearly distinguish known activators of the two pathways. The luminescence of the two reactions occurs at different wavelengths, and hence we were able to simultaneously assess the activation of the two signaling pathways in the same sample with either cell lysates (Figure 4.4g) or intact HEK293T cells (Figure E4.10c-e) by providing both substrates together and monitoring luminescence at different wavelengths.

4.4 Conclusion

Computational enzyme design to date has been constrained by the number of available scaffolds, which limits the extent to which catalytic configurations and enzyme-substrate shape complementarity can be achieved [14, 15, 16]. The use of deep learning to generate large numbers of de novo designed scaffolds here eliminates this restriction; moving forward, the more accurate RoseTTAfold [38] and AlphaFold2 [31] should enable still more effective protein scaffold generation [18, 39] through family-wide hallucination and other generative approaches. The diversity of scaffold pocket shapes and sizes enabled the exploration of a range of catalytic geometries and the maximization of substrate-enzyme shape complementarity; to our knowledge, no native luciferases have folds similar to LuxSit, and the enzyme has high specificity for a fully synthetic

luciferin substrate that does not exist in nature. With the incorporation of 2-3 substitutions that provide a more complementary pocket to stabilize the transition state, LuxSit-i has higher activity than any previously de novo designed enzyme; the $k_{\text{cat}}/K_{\text{M}}$ of $106 \text{ M}^{-1}\text{s}^{-1}$ is in the range of native luciferases. This is a notable advance for computational enzyme design, as tens of rounds of directed evolution were required to obtain catalytic proficiencies in this range for a designed retroaldolase, and the structure was remodeled considerably [40]; in contrast, the predicted differences in ligand-sidechain interactions between LuxSit and LuxSit-i are very subtle (Figure S4.1b; achieving such high activities directly from the computer remains an outstanding goal for computational enzyme design). The small size of LuxSit-i makes it well-suited as a genetic tag for capacity-limited viral vectors, biosensor development, and fusions to proteins of interest. On the basic science side, the small size, simplicity, and high activity make LuxSit-i an excellent model system for computational and experimental studies aimed at improving understanding of the luciferase catalytic mechanism. Extension of the approach used here to create similarly specific new luciferases for synthetic luciferin substrates beyond DTZ and h-CTZ would considerably extend the multiplexing opportunities illustrated in Figure 4.4 (particularly with the recent advances in microscopy) [41], leading to multiplexed luminescent toolkits. More generally, our family-wide hallucination method opens up an almost unlimited number of new scaffold possibilities for substrate binding and catalytic residue placement, which is particularly important in cases where the reaction mechanism and how to promote it are not completely understood – many structural and catalytic hypotheses can be readily enumerated with different catalytic residue placements in shape and chemically complementary binding pockets. While luciferases are unique in catalyzing the emission of light, the chemical transformation of substrates into products is common to all enzymes, and the approach developed here should be readily extendable to a wide variety of chemical reactions.

4.5 Methods

4.5.1 Materials and general methods

Synthetic genes and oligonucleotides were purchased from Integrated DNA Technologies or GenScript. The synthetic gene was inserted into modified pET29b+ vectors, containing N- or C-terminal hexahistidine tag. Restriction endonucleases, Q5 PCR polymerase, USER enzyme, NEB-Next end repair module, and T4 ligase were purchased from NEB. Plasmid DNA, PCR products, or digested fragments were purified by Qiagen DNA purification kits. DNA sequences were analyzed by Genewiz. Coelenterazine (CTZ) was purchased from Gold Biotechnology. Diphenylterazine (DTZ), pyridyl diphenylterazine (8pyDTZ), and Furimazine (FRZ) were purchased from MedChemExpress. All other coelenterazine analogs (bis-CTZ: bisdeoxycoelenterazine; f-CTZ: f-Coelenterazine; e-CTZ: e-Coelenterazine-F; PP-CTZ: methoxy e-Coelenterazine; v-CTZ: v-Coelenterazine; h-CTZ: 2-deoxycoelenterazine) were ordered from NanoLight. All other chemicals were purchased from Sigma-Aldrich or Fisher Scientific and used without further purification. To identify the molecular mass of each protein, intact mass spectra were obtained via reverse-phase LC/MS on an Agilent 6230B TOF on an AdvanceBio RP-Desalting column and subsequently deconvoluted by Bioconfirm software (B.07.00) using a total entropy algorithm. ÄKTA pure M with UNICORN 6.3.2 Workstation control (GE Healthcare) coupled with a Superdex 75 Increase 10/300 GL column was used for size exclusion chromatography. DNA and protein concentrations were determined by an 8-channel NanoDrop UV/vis spectrometer. CD spectra and CD melting experiments were performed by default setting on a J-1500 Circular Dichroism Spectropolarimeter (Jasco). All luminescence measurements were acquired by a Biotek Synergy Neo2 Multi-Mode Plate Reader. To convert relative arbitrary unit (RLU) to the number of photons, Neo2 plate reader was calibrated by determining the chemiluminescence of luminol with known quantum yield in the presence of horseradish peroxidase and hydrogen peroxide in K₂CO₃ aqueous solution as previously described [42]. SDS PAGE and luminescence images were captured by a

Bio-Rad ChemiDoc XRS+. HEK293T (CRL-11268) and HeLa (CCL-2) cells were obtained from ATCC. Images were analyzed using the Fiji image analysis software.

4.5.2 General procedures for protein production and purification

Lemo21(DE3) strain was used for transformation with the pET29b+ plasmid encoding the gene of interest. Transformed cells were grown for 12 h in LB medium supplemented with kanamycin. Cells were inoculated at 1:50 ratio in 100 mL fresh TB medium, grown at 37 °C for 4 h, and then induced by IPTG for an additional 18 h at 16 °C. Cells were harvested by centrifugation at 4,000g for 10 min and resuspended in 30 mL lysis buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 30 mM imidazole, and Pierce™ Protease Inhibitor Tablets). Cell resuspensions were lysed by sonication for 5 min (10 s per cycle). Lysates were clarified by centrifugation at 24,000g at 12 °C for 40 min and pre-equilibrated with 1 mL of Ni-NTA nickel agarose at 4 °C for 1 h. The resin was washed twice with 10 mL wash buffer and then eluted in 1 mL elution buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM imidazole). The eluted proteins were purified by size exclusion chromatography in PBS. Fractions were collected based on A280 trace, snap-frozen in liquid nitrogen, and stored at -80 °C.

4.5.3 Computational design of idealized scaffolds

Our generation of idealized NTF2-scaffolds can be divided into four parts: Generation of seed-structures, optimization of backbone geometries using trRosetta-based hallucination, generation of structure-conditioned sequence models to bias design, and design and filtering.

Generation of seed structures

We thought to increase the set of NTF2 structures by complementing experimentally resolved structures from the PDB with highly accurate models generated by trRosetta [43]. To achieve this, we first collected 85 NTF2-like protein structures from the PDB based on SCOPe annotation (d. 17.4 SCOPe v2.05). Corresponding sequences were then used as queries to collect

sequence homologs from UniProt by performing 8 iterations of hhblits at $1e-20$ e-value cutoff against unclust30_2018_08 database; default filtering cutoffs were relieved (-maxfilt 10000000 -neffmax 20 -nodiff -realign_max 10000000) to maximize the number of the output hits. All the hits were redundancy reduced using cd-hit [44] with a sequence identity cutoff of 60% yielding a set of 7,573 candidates for modeling.

To generate inputs for structure modeling with trRosetta, we built multiple sequence alignments (MSAs) for each of the 7,573 selected sequences with hhblits using a more conservative e-value cutoff of $1e-50$; the resulting MSAs were also complemented by hits from hmmsearch against uniref100 (release-2019_11) with the bit-score threshold of 115 (i.e. 1bit per position). After joining the above two sets of alignments and filtering them at 90% sequence identity and 75% coverage cutoffs, only sequences with more than 50 homologs in the corresponding MSAs were retained for modeling (2,005 sequences). The filtered MSAs along with information on the top 25 putative structural homologs as identified by hhsearch against the PDB100 database of templates were used as inputs to the template-aware version of trRosetta [45] to predict residue pair distances and orientations. Network predictions were then used to reconstruct full atom 3D structure models using a Rosetta-based folding protocol described previously [20].

Hallucination of idealized NTF2s

Seeking to idealize the native structure seeds, we reasoned that trRosetta, a convolutional residual neural network, which predicts residue-residue orientations and distances from sequence, could serve as a key component in a protein idealizer. Previously, this network has been used to generate diverse proteins that resemble the “ideal” structures of de novo designed proteins by changing the protein sequence to optimize the contrast (KL-divergence) between the predicted geometry and that of randomly generated sequences [17].

For our purpose, the desired fold-space is not diverse but instead focused on the NTF2-like topol-

ogy. To guarantee generation of ideal structures within this fold-space, we implemented a new fold-specific loss-function, which biased hallucinations based on observed geometries in native crystal structures. As many experimentally characterized NTF2s contain non-ideal regions, we began by creating a set (χ) of trimmed but ideal NTF2s by manually removing non-ideal structural elements such as kinked helices, and long or rarely observed loops. For each seed structure, we then used a structure-based sequence alignment method to find equivalent positions between the seed structure and χ . Residue pairs were considered to be in a conserved tertiary motif (TERM) if there were 5 or more equivalent positions in χ . The smooth probability distributions based on observed geometries in χ were then computed. For distances we used a Gaussian distribution with mean equal to the true distance denoted by D and standard deviation denoted by σ equal to 0.5 \AA . The probability density function for distances d is given by:

$$f(d; D, \sigma) = \frac{1}{\sqrt{(2\pi\sigma^2)}\exp\left(-\frac{(d-D)^2}{2\sigma^2}\right)}$$

Using this density function one can construct a categorical distribution for binned distances by evaluating this function at the centers of the bins and then normalizing by a sum of all values in different bins. Similarly, a von Mises distribution was used for omega angle smoothing with probability density function given by $(\omega - \Omega)$ where $N(\kappa)$ is a normalizing constant, Ω is the crystal value, κ is the inverse variance chosen to be 100, and ω is the smoothed angle. For phi and theta angles a von Mises-Fisher blur is given by $f(x; \mu, \kappa) = N(\kappa)\exp[\kappa\mu^T x]$ where $N(\kappa)$ is a normalizing constant, μ is a unit vector on a 3D sphere corresponding to the phi and theta angles from the crystal structure, x is a smoothed unit vector, and κ is the inverse variance chosen to be 100.

Next, we converted those probability distributions to energy landscapes (i.e. - negative log likelihoods) and sought to minimize the expected energy. This soft restraint encouraged the network to seek out the consensus structure, while still allowing deviations where needed. Specifically, we formulated the fold-specific loss as:

$$L_{fold} = \sum_{x \in \{d, \omega, \theta, \phi\}} \left[\sum_{i,j=1}^L \sum_{k=1}^{N_x} -m_{ij} p_{x,ijk} \ln(s_{x,ijk}) \right] / \sum_{i,j=1}^L m_{ij}$$

$$m_{ij} = \{1 \text{ if } i \text{ and } j \text{ are in a TERM ; else } 0\}$$

where p is the network prediction and s is the smoothed probability distribution of the conserved residue pairs. For the second part of the loss function and similar to previous work [46], we sought to maximize the Kullback–Leibler (KL) divergence between the predicted probability distribution and a background distribution for all i, j residue pairs not in a TERM.

$$L_{hall} = - \sum_{x \in \{d, \omega, \theta, \phi\}} \left[\sum_{i,j=1}^L \sum_{k=1}^{N_x} (1 - m_{ij}) p_{x,ijk} \ln(p_{x,ijk}/b_{x,ijk}) \right] / \sum_{i,j=1}^L (1 - m_{ij})$$

where b is the background distribution and N_x is the number of bins in each probability distribution ($N_d = 37$, $N_{\omega, \theta} = 25$, $N_{\phi} = 13$). Briefly, b is calculated by a network of similar architecture to trRosetta trained on the same training data, except it is never given sequence information as an input. The final loss is given by:

$$L = L_{fold} + L_{hall}$$

We used a Markov Chain Monte Carlo (MCMC) procedure to search for sequences that trRosetta predicted to fold into structures that minimize this loss function. We allowed four types of moves with different sampling probabilities: mutations ($p=0.55$), insertions ($p=0.15$), deletions ($p=0.15$), and moving segments ($p=0.15$). Mutations randomly changed one amino acid to another, with an equal transition probability for all 20 amino acids. Insertions inserted a new amino acid (all equally likely) into a random location subject to the KL-divergence loss. Deletions deleted a random residue from the same locations. Finally, we also allowed “segments” to move, cutting and pasting themselves from one part of the sequence to another, while maintaining the same overall segment order. Here, a “segment” is a continuous stretch of amino acids all subject to fold specific loss, often composed of a single strand or helix. Starting from a random sequence of an initial length

(typically 120 amino acids), we used the standard Metropolis criteria to accept or reject moves:

$$A_i = \min[1, \exp(-(L_i - L_{i-1})/T)]$$

where A_i is the chance of accepting the move at step i , L_i is the loss at the current step, L_{i-1} is the loss at the previous step and T is the temperature. The temperature started at 0.2 and was reduced by half every 5k steps. Generally, it took 30k steps to converge.

Structure-conditioned multiple sequence alignment

Given the complexity of the NTF2-like protein fold, we hypothesized that it was necessary to impose sequence design rules to disfavor alternative states (negative design). Towards this end, we computed a structure-conditioned multiple sequence alignment based on native NTF2-like proteins. Specifically, we used TMalign [47] to superimpose each of the 2005 predicted native structures onto each hallucinated backbone. Next, to find structurally corresponding positions, we implemented a structure-based dynamic programming algorithm, similar to the Needleman-Wunsch algorithm [48]. However, instead of using the amino acid similarity as the scoring metric, we used a tunable structure-based score function. After aligning the two structures, we scored the structural similarity of any two residues by empirically weighting several metrics: (1) Distance between Ca atoms, (2) differences between backbone torsion angles (phi and psi) backbone torsion angles, and (3) the angle (degrees) between the vectors pointing from Ca to Cb in each residue. To calculate the unweighted score for each component, we normalized each by a maximum possible value (180 degrees for angles and 10Å for distances) and included a “set point” that was approximately delineated when we judged a metric to indicate two residues to be more similar than not. Values above this setpoint are positive, indicating two residues are similar and values below the set point indicated two residues are dissimilar.

$$Score_{unweighted} = (set_{point} - value)/max_{value}$$

Each value was scaled by its normalized weight and summed to give an overall similarity score between any two amino acids.

Score Component	Weight	Setpoint	Max value
Phi	1.0	45°	180°
Psi	1.0	45°	180°
CaCb angle	2.0	30°	180°
CaCa distance	2.5	3Å	10Å

These similarity scores were used as the similarity metric in our dynamic programming algorithm, in place of the typical BLOSUM62 similarity metric. We used a gap penalty of 0.1 and an extension penalty of 0.0. Finally, after concatenating all the structure-conditioned aligned sequences, we used PSI-BLAST-exB [49, 50] to compute sequence redundancy weighted log-odds scores for each amino acid at each position (position-specific scoring matrices, PSSMs).

Sequence design

To design the resulting backbones, we sought, in addition to the sequence patterns captured in the PSSM, to further specify the backbone conformation and functionalize the pocket, by installing entire hydrogen bonding networks from native NTF2-like proteins. We compiled two sets of hydrogen bonding networks: a set for the cavity containing 85 networks and another set of networks connecting the C-terminal region of the first helix with the third beta-strand containing 25 networks. In 20 independent attempts for each backbone, we randomly grafted a network from each set, fixed the identities of hydrogen bonding residues, and designed the sequences for all other positions under PSSM constraints. The resulting models were filtered for various backbone quality metrics and for maintenance of hydrogen bonding networks in the absence of constraints, resulting in a total of 1615 idealized scaffolds.

4.5.4 RIFdock tuning files

The hierarchical search framework of RifDock is a powerful way to search through 6-dimensional rigid body orientations. While originally designed to work with physics-based forcefields, the scoring machinery can easily be modified to do other things. A system was added called “Tuning Files” that allows one to tune the energetics of rifdock by “requiring” specific interactions. Specified interactions can range from specific hydrogen bonds, to specific bidentates, and even to specific hydrophobic interactions. The specifics are that during the RifGen stage, each stored rotamer is compared against a list of definitions in the Tuning File. If the rotamer satisfies a definition, it is stored into the RIF with a “Requirement Number”. Later during RifDock, these Requirement Numbers are available during scoring and the presence or absence of certain rotameric interactions may be used to penalize or even completely discard dock solutions. In this work, the Tuning Files were used to require the specific hydrogen bond interactions between the arginine and the secondary amine in the pyrazine ring of the colenterazine-like substrate. The documentation for RifDock tuning files can be found [at the link here](#).

4.5.5 Designing theozyme architectures into de novo NTF2 scaffolds

De novo design of luciferases can be divided into three main steps – scaffold construction, substrate placement with required interactions, and sequence design. With the idealized NTF2-like scaffolds in hand, we manually selected 5 diverse anionic DTZ conformers and used the Rotamer Interaction Field (RIF) docking method [29] to exhaustively search a large space of interacting side chains to the anionic form of DTZ. Chemically, deprotonation of N1 hydrogen is the first step to forming an anionic species (Figure E4.1). We first generated RIF using Rifgen [29] to sample the placement of amino acid sidechains around DTZ. We required the placement of a positively charged Arginine sidechain by a tuning file (see below) to stabilize the formation of negatively charged N1 atom where the deprotonation initially occurs and enumerated large numbers of possible sidechain interactions with the rest of DTZ. As a comparison, we also placed tryptophan or

histidine next to the N1 atom in another tuning file.

```
HBOND_DEFINITION
```

```
N1 1 ARG
```

```
END_HBOND_DEFINITION
```

```
REQUIREMENT_DEFINITION
```

```
1 HBOND N1 1
```

```
END_REQUIREMENT_DEFINITION
```

Rifdock was then used to hierarchically search for the best combination of RIF to place on the input backbone. Although the negative charge can move to another electronegative atom O1 via resonance of the imidazopyrazinone core, it is unclear which anionic species is more critical for the luciferase-catalyzed luminescence emission. Thus, we let RIFdock place the polar rotamers on the basis of hydrogen-bond geometry to O1 and apolar rotamers to DTZ without specific requirements. In the next docking step, we parsed the `-scaffold_res` argument with a list of residue numbers as scaffold backbone positions that were annotated as pocket residues to allow a hierarchical search of RIF placement. We allowed the RIF placements in the pocket residues including or excluding pre-defined hydrogen bond networks. After RIFdock, we continued for Rosetta sequence design where the score function was reweighted for higher buried_unsat_penalty [51], and the amino acid selection was biased by giving a pre-generated PSSM file via SeqprofConsensus task operation. This would minimize buried unsatisfied residues and increase pre-organized architectures in the core that are known to be beneficial for a catalytic pocket [52]. Two rounds of Rosetta FastDesign calculation were included: we restricted the RIF residues and core HBNets to repacking in the first round while we allowed the re-design of other residues based on PSSM during the Monte Carlo simulated annealing procedure. After the surrounding residues were optimized to retain the RIF interactions, we enabled the re-design of RIF residues, giving Rosetta a chance

to find efficient aromatic and hydrophobic packing around DTZ while catalytic residues (the N1 requirement) were still limited to only repacking. The final set of designs was obtained after filtering by Rosetta ligand-binding interface energy, shape complementarity, contact molecular surface, number of HbondsToResidue, and the presence of N1 hydrogen bond.

4.5.6 Structure prediction of LuxSit with AlphaFold2 and comparison to design model

To computationally assess the accuracy of our LuxSit design model, we performed single sequence structure prediction using AlphaFold2. All models were run with 12 recycles and generated models were relaxed using AMBER13. The model with the highest pLDDT was used for comparison to the Rosetta design model and structural superpositions were performed using the Theseus alignment tool to determine backbone RMSD between the design model and AlphaFold2 model [31].

4.5.7 Computational design and characterization of de novo luciferases for h-CTZ

To customize a shape complementarity catalytic pocket that can accommodate and catalyze chemiluminescence of another structurally distinct luciferin substrate (2-deoxycoelenterazine, h-CTZ), we sought to use a more diverse set of scaffolds. We first used a deep-learning based protein sequence design method, ProteinMPNN [32] to redesign the whole sequences of the hallucinated NTF2 scaffolds described earlier and the de novo NTF2-like superfamily reported previously [53]. Next, the protein structures of all resulting ProteinMPNN sequences were predicted by AlphaFold2 [31]. 6234 scaffolds with diverse pocket geometries were obtained by filtering the pLDDT score greater than 92. With these scaffolds in hand, we selected three different h-CTZ conformers and used the RIFdock design strategy described above to search for the sidechain rotamer placements in these scaffolds. Since we've learned from LuxSit design that the N1-Arg and O1-His interactions are critical for catalyzing luminescence emission, both interactions were set as requirements

in a tuning file (see an example below) to ensure all Rifdock outputs have N1-Arg and O1-His interactions.

```
HBOND_DEFINITION
```

```
N1 1 ARG
```

```
O1 1 HIS
```

```
END_HBOND_DEFINITION
```

```
REQUIREMENT_DEFINITION
```

```
1 HBOND N1 1
```

```
2 HBOND O1 1
```

```
END_REQUIREMENT_DEFINITION
```

At this stage, we generated 215k RIFdock outputs in which we subsequently fixed the N1-Arg and O1-His interactions (by applying atom constraints) and allowed Rosetta to redesign all residues within 4 Å of the ligand. The resulting Rosetta designs were prefiltered by contact molecular surface (> 350), Rosetta ddG (< -50), and the presence of N1-Arg and O1-His interactions. The prefiltered Rosetta sequences (30k) were then optimized by ProteinMPNN while all amino acid identities within 4 Å of the ligand were kept fixed. All ProteinMPNN sequences were predicted by AlphaFold2 to obtain predicted 3D protein models where we evaluated the pLDDT score (> 85), Ca RMSD (< 1.2 Å) to the corresponding Rosetta model, and the numbers of hydrogen bonds to both hypothetical catalytic Arg (> 2) and His (> 1) residues for preorganization. Finally, 46 sequences passed the filters, and we ordered them as eBlocks gene fragments for experimental characterizations.

Each synthetic gene was inserted into a modified pET29b vector between two BsaI sites (Golden gate assembly) and transformed into BL21 competent *E. coli*. The cells were inoculated in LB

and grew in a 96-deep well plate with 1mM IPTG at 37 °C for 16h. For the luciferase activity screening, the cells were harvested by centrifugation at 3,800g for 5 min and the pellets were lysed by the BugBuster reagent. Cell lysates were collected by centrifugation at 4000g for 20 min. The His-tag proteins were captured by nickel magnetic beads from cell lysates and bound proteins were eluted in elution buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM imidazole). The protein concentrations were determined by Bradford assay and the activity of each luciferase was evaluated individually in the presence of 1 μ M purified protein and 25 μ M h-CTZ in PBS. Through this process, we identified two designs (HTZ3-D2 and HTZ3-G4) that showed luciferase activity and substrate selectivity to h-CTZ. We scaled up the protein expression by the general procedure described above and the purified proteins were used for the characterization shown in Figure E4.4. Serially diluted h-CTZ was mixed with 500 nM HTZ3-D2 or HTZ3-G4 in PBS and the concentration-dependent luminescence was recorded for 30 mins (0.1 s integration and measurements were taken every 1 min). All data points were plotted as the average of the first 10 mins light output and fitted to Michaelis-Menten equation.

4.5.8 Computational SSM to estimate mutation binding free energy

Rosetta cartesian_ddg application [53, 54] was used to computationally estimate enzyme and substrate binding free energy. The LuxSit design model was relaxed beforehand in cartesian space with the substrate-bound. For the 21 positions that were experimentally screened for single mutation effects on luciferase activity, each residue was computationally mutated into other amino acid types. Packing and cartesian relaxation were subsequently performed to evaluate the final score in REU. This procedure was applied three times in parallel for both substrate-bound and apo-states. The average of the three calculation results was used to calculate the relative binding free energy (ddG_{bind}) by subtracting the total score of the apo-state from the complex state. The change of binding free energy upon mutation was plotted against experimental activity for every position in Figure ???. The wild-type system shown in red dots underwent the same amount of computation by keeping the residue type but applying sidechain repacking and relaxation. The ddG_{bind} rank of the

wild type among the mutations is shown with a colored heat map in Figure ??.

4.5.9 Construction and screening of designed luciferase libraries

The construction of assembled gene libraries was described previously in detail [46]. In brief, the amino acid sequences of all designed luciferases were first reverse-translated into *E. coli* codon-optimized DNA sequences. All DNA sequences were categorized into multiple sub-pools by the gene length (500 designs per sub-pool). Each gene was subsequently split into two fragments (fragment A and fragment B) and added outer and inner primer sequences to the 5' and 3' end (e.g., Outer_oligoA_5primer + design_half_A + Inner_oligoA_3primer and Inner_oligoB_5primer + design_half_B + Outer_oligoB_3primer). All oligos were ordered in one Twist 250nt Oligo Pool. To construct the library of each sub-pool, polymerase chain reaction (PCR) with oligoA_5primer/oligoA_3primer or oligoB_5primer/oligoB_3primer oligonucleotide pairs was used to amplify the individual fragment A or fragment B from each sub-pool. The pool-specific sequences were removed with Uracil Specific Excision Reagent (USER) followed by NEB End Repair kit. Outer primers (oligoA_5primer and oligoB_3primer) were then used for fragment A and fragment B assembly and amplification. The assembled full-length fragment was digested with XhoI/HindIII and ligated into a predigested pBAD/His B vector. All ligation products were used to transform ElectroMAX™ DH10B Cells, which were next plated on 150 mm x 15 mm LB agar plates supplemented with carbenicillin and L-arabinose. We sequenced 30 random colonies and 11 of the sequences were in our designed library. The plates (2000 colonies per plate) were incubated at 37 °C overnight to form bacterial colonies and left at 4 °C for another 24 h. To directly image luminescence activity from bacterial colonies, we sprayed the PBS solution containing 30 μM DTZ to each agar plate, waited for 2 min, and the luminescence images were acquired and processed with Bio-Rad ChemiDoc XRS+. After screening 15 plates, active colonies were collected for sequencing, protein expression, and other downstream characterization where LuxSit was selected from three active designs showing catalytic signals above the background.

4.5.10 Construction and evaluation of LuxSit site saturation mutagenesis libraries

To create libraries of each single amino acid substitution at residues 13, 14, 17, 18, 35, 37, 38, 49, 52, 53, 56, 60, 65, 81, 83, 94, 96, 98, 100, 110, and 112, forward oligos mixture with degenerate codons (NDT, VHG, and TGG = 1:1:0.1 ratio) and an overlapped reverse oligo were used to amplify the plasmid of LuxSit. The resulting PCR products were circularized by Gibson Assembly protocol and were subsequently used to transform ElectroMAXTM DH10B Cells. The cells were plated on 150 mm x 15 mm LB agar plates supplemented with carbenicillin and L-arabinose, incubated at 37 °C overnight, and left at 4 °C for another 24 h. As described in the screening of luciferase libraries, colony-based screening by spraying DTZ solution was used to identify active colonies. Inactive colonies were also randomly picked. As a result, a total of 32 colonies were picked for each residue library. 32 x 21 individual colonies were grown in 1 mL of TB supplemented with carbenicillin and L-arabinose in 96-well deep-well culture plates. The plates were shaken at 37 °C overnight (16-18 h) on 96-well plate shakers at 1,100 rpm. Cells were pelleted by centrifugation at 3,800g for 15 min in a tabletop centrifuge. Media was discarded and the cell pellets were resuspended in 0.2 mL BugBuster HT Protein Extraction buffer. The plates were transferred back to 96-well plate shakers and incubated at 1,100 rpm for an additional 30 min. Cellular debris was pelleted again by centrifugation at 4,000g for 15 min, soluble lysates were transferred to a new semi-deep 96-well plate, and incubated with 10 μ L of magnetic Ni-NTA beads for 30 min to allow binding. The magnetic extractor was used to first transfer the beads from the binding plates to wash plates with 200 μ L IMAC wash buffer in each well, and then transfer the beads to elution plates containing 30 μ L IMAC elution buffer in each well. The concentrations of all proteins in each well were determined by the Bradford assay directly. The elution solution in each well was used to make a 25 μ L protein solution at indicated concentration and mixed with 25 μ L of 50 μ M DTZ PBS solution. The luminescence signals were acquired over a course of 15 min while the actual point mutation was identified by sequencing. Thus, the mutation-to-activity relationship can

be mapped. To evaluate whether these beneficial mutations are synergistic, we ordered individual mutants with combinatorial mutations at residue 14, 60, 96, 98, and 110, expressed, and purified these LuxSit variants for kinetic, emission spectra, and luminescence intensity. We identified four mutants that can produce 47 to 77-fold more photons than the parent LuxSit. We assigned one of which, LuxSit-f (A96M/M110V), for its strong initial flash emission. Since the mutations at residue 96 and 110 are robust and mutations at residue 60 are versatile, we generated a fully randomized library at 60, 96, and 110 positions to exhaustively screen all possible combinations. After the colony-based screening, we identified many colonies with strong luciferase activities with DTZ (Figure E4.6). Among all selected mutants, Arg60 is confirmed to be mutable, Ala96 prefers larger hydrophobic sidechains (Leu, Ile, Met, and Cys), and Met110 favors hydrophobic residues (Val, Ile, and Ala). A newly discovered mutant R60S/A96L/M110V with more than 100-fold higher photon flux over LuxSit was assigned LuxSit-i for its high brightness.

4.5.11 In vitro characterization of photoluminescence properties

For Michaelis–Menten kinetics measurements, 25 μL of serial diluted DTZ substrate in Tris pH 8.0 buffer was added into the wells of a white 96-well half-area microplate containing 25 μL of purified luciferases (final enzyme concentration: 100 nM; substrate concentration: 0.78 to 50 μM). Measurements were taken every 1 min (0.1 s integration and 10 s shaking between each interval) for a total of 20 min. Initial velocities were estimated as the average of the light intensities from the first three data points to fit the Michaelis-Menten equation. All relative arbitrary unit (RLU) per second values were converted to photon/s by the luminol-H₂O₂-HRP calibration method [42]. Following the equation: $I_{\text{max}} = \text{LQY} \times k_{\text{cat}} \times [\text{E}]$, I_{max} is the maximal photon flux (photon s⁻¹), $[\text{E}]$ is the total enzyme concentration, and V_{max} is the maximum photon flux per molecule (photon s⁻¹ molecule⁻¹) from the fitting of the Michaelis-Menten equation. To determine the luminescent quantum yields, 125 pmol of individual substrate in 25 μL PBS was injected into 25 μL PBS containing 100 nM corresponding luciferase. DTZ was used for all LuxSit variants while CTZ was used as the substrate of native RLuc. The luminescence signals were monitored until the

reactions were completed (0.1 s integration and measurements were taken every 5 s for a total of 40 min). The sum of luminescence photon counts was normalized to the total photon counts of RLuc/CTZ pair (LQY = $5.3 \pm 0.1\%$)²⁰ to derive relative luminescent quantum yields of LuxSit variants (Figure E4.7c). k_{cat} values for each individual enzyme were calculated using the equation: $k_{\text{cat}} = v_{\text{max}} / \text{LQY}$. To record emission spectra, 25 μL of 50 μM DTZ in PBS was injected into 25 μL of 200 nM pure luciferases and the emission spectra were collected with 0.1 s integration and 2 nm increments from 300 to 700 nm. In vitro luminescence activity measurements of LuxSit-i expressing HEK293T or HeLa cells were done similarly as 15,000 intact cells or lysates were used in the assay instead of purified luciferases. To evaluate the substrate specificity, 25 μL of 50 μM substrate analogs in PBS was added to 25 μL of 200 nM indicated luciferases, and the signals were recorded over 20 min. Data were shown as the total luminescence signal over the first 10 min. We normalized the data by setting the highest emission substrate at 100%. The unnormalized plot of Figure 4.4c was attached as Figure S4.3.

4.5.12 Circular dichroism (CD)

Purified protein samples were prepared at 15 μM in pH 7.4 10 mM phosphate buffer. Spectra from 190 nm to 260 nm were recorded at 25 °C, 50 °C, 75 °C, 95 °C, and after cooling back to 25 °C. Thermal denaturation was monitored at 220 nm from 25 °C to 95 °C (1 °C per min increments). T_m values were not reported because no obvious inflection points of the melting curves.

4.5.13 Mammalian cell culture and transfection

HEK293T and HeLa cell lines were maintained at 37 °C with humidified 5% CO₂ atmosphere and cultured in Dulbecco's Modified Eagle's Medium (DMEM, GIBCO) supplemented with 10% fetal bovine serum (FBS, Sigma). Cells were transfected with Turbofectin 8.0 (Origene) with 500 μg of plasmid DNA. After 24 h at 37 °C in a CO₂ incubator, the medium was removed, and cells were collected and resuspended in Dulbecco's phosphate-buffered saline (DPBS).

4.5.14 Fluorescence Microscopy and image analysis

Cells were washed twice with HBSS and subsequently imaged in HBSS in the dark at 37 °C. Right before imaging, cells were incubated with 25 μ M DTZ. Epifluorescence imaging was conducted on a Yokogawa CSU-X1 microscope equipped with a Hamamatsu ORCA-Fusion scientific CMOS camera and Lumencor Celesta light engine. Objectives used were: 10x, NA 0.45, WD 4.0 mm, 20x, NA 1.4, WD 0.13 mm, and 40x, NA 0.95, WD 0.17–0.25 mm with correction collar for cover glass thickness (0.11 mm to 0.23 mm) (Plan Apochromat Lambda). Imaging for BFP utilized a 408 nm laser, 432/36 nm dichroic, and a 440/40 nm emission filter (Semrock). Exposure times were 200 ms for BFP and 10 s for luminescence. All epifluorescence experiments were subsequently analyzed using NIS Elements 5.30 software.

4.5.15 Multiplex dual-luciferase reporter assay for the cAMP/PKA and NF- κ B pathways

HEK293T cells were grown in a tissue culture-grade white 96-well plate and transfected with indicated CRE-RLuc, NF κ B-LuxSit-i, and CMV-CyOFP plasmids. 24 h after transfection, the medium was replaced by 2 μ M of Forskolin (FSK) or 300 ng/mL human tumor necrosis factor alpha (TNF α) in regular cell media. 23 h after stimulation, the cells were resuspended in DPBS by pipette mixing. 25 μ L of DPBS containing 30,000 intact cells was mixed with 25 μ L of CellLytic M for 15 min to make cell lysates. For intact cell assay, 25 μ L of DPBS containing 15,000 intact cells was mixed with 25 μ L of PP-CTZ (2 μ M) or/and DTZ (10 μ M) in DPBS. For cell lysate assay, 25 μ L of cell lysate was added to 25 μ L of PP-CTZ (2 μ M) or/and DTZ (10 μ M) to initiate luminescence reactions. The signals were recorded every 1 min for a total of 10 min. The light signals were collected in the substrate-resolved mode without filters and with 528/20 and 390/35 filters under the spectrally resolved mode. Area scanning of the CyOFP fluorescence intensity at 480 nm (excitation wavelength) and 580 nm (emission wavelength) was used to estimate the total cell numbers and transfection efficiency. The reported unit was the average of the first 10 min

luminescence (RLU) over the relative fluorescence units (a.u.). To derive fold-of-activation, all values were normalized to the corresponding non-stimulated control.

4.5.16 Quantum mechanics calculation for the energy profile of DTZ luminescence

All calculations were carried out by using the Gaussian 16 Rev. A 03. Given the benchmark study done by Martin Head-Gorden and co-workers recently [55], geometry optimizations and frequency calculations were performed at the (u) ω B97X-D/6-31+G(d) level of theory [56, 57]. Frequency calculations were conducted at the same level of theory to confirm the presence of local minima (no imaginary frequencies) and transition states (one imaginary frequency) on the PES. Subsequent single-point energies were computed at the (u) ω B97X-D/def2TZVP level. Solvent effect was modeled by employing the CPCM model. Conformational searches were conducted using the CREST conformer-rotamer ensemble sampling tool version 2.10.2 with xtb version 6.3.3 to ensure the substrate showed are lowest energy conformers [58, 59]. Intrinsic reaction coordinate (IRC) calculations were performed to verify that the saddle points found were true TSs connecting the reactants and the products. TDDFT result were given by using the same level and method with DFT part, setting nstates=10 to ensure the calculation accuracy. All thermodynamic quantities (1 mol/L, 298.15 K) were computed in the GoodVibes code [60] with quasiharmonic corrections [61]. 3D renderings of stationary points were generated using open-source PyMOL.

4.5.17 Molecular dynamics simulations of proposed enzyme-ligand complexes

Molecular dynamics simulations were performed using the GPU code (pmemd) [62] of the AMBER 16 software package. Substrate structures were optimized at the B3LYP-D3/6-31G(d) level of theory [63, 64, 65] using Gaussian16, Revision A.03. Rosetta parameter files for the substrates were generated using the provided molfile2params.py script on the optimized structures. Substrate starting positions were determined using the ligand_dock application [66, 67, 68] of the Rosetta3

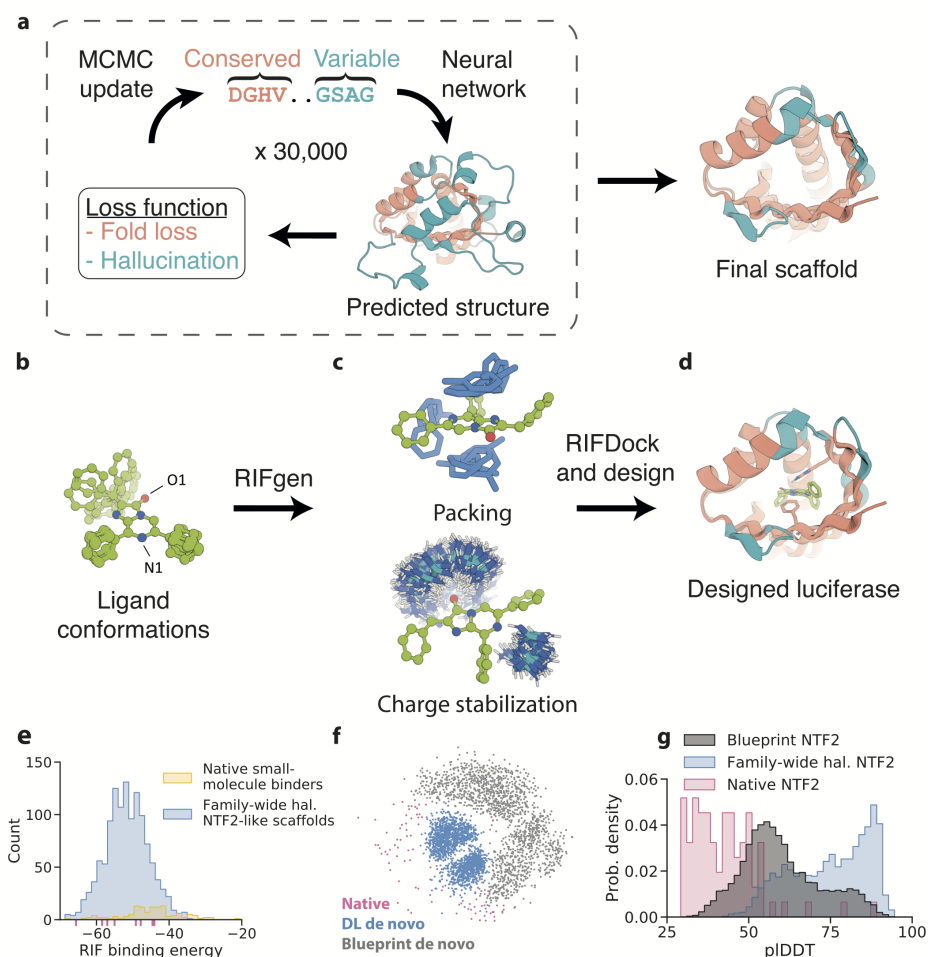
software package. Each substrate was docked using 500 trajectories starting from the original design position. Docked structures with the lowest `interface_delta` and `ligand_auto_rms_no_super` were chosen for simulation starting coordinates. AMBER parameter files for the substrates were generated using the general AMBER force field (GAFF) [69]. Partial charges were set to fit the electrostatic potential generated at the previously mentioned level of theory using the RESP model [70]. Charges were calculated using the Merz-Singh-Kollman scheme [71, 72] using Gaussian16, Revision A.03. The system was immersed in a pre-equilibrated octahedral box with a 10 Å buffer of TIP3P water molecules [73] using the `tleap` module. Explicit Na⁺ and Cl⁻ ions were added to neutralize the total charge of the system. All subsequent calculations were performed using the AMBER 14 force field (`ff14sb`) [74]. Two minimization steps were conducted in serial, each consisting of 2,500 steepest descent steps and 2,500 conjugate gradient steps. Protein and substrate atoms were restrained with a force constant of 500.0 kcal mol⁻¹ Å⁻² in the first minimization step, allowing the solvent to minimize. Sidechain atoms were unrestrained in the second minimization, but backbone and substrate atoms were restrained with a force constant of 50.0 kcal mol⁻¹ Å⁻². Minimizations were followed by two equilibration steps where backbone and substrate atoms were restrained with a force constant of 30.0 kcal mol⁻¹ Å⁻². In the first equilibration step, the system was heated from 0-300 K over 300 ps using constant-volume and periodic-boundary conditions and a 1 fs time step. The SHAKE algorithm was used to constrain bonds involving hydrogens. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method [75] and an 8 Å cutoff was applied to the Lennard-Jones and electrostatic interactions. A second equilibration was then conducted for 50 ns using a 2 fs time step at constant pressure using isotropic position scaling with the Monte Carlo barostat and a pressure relaxation time of 5.0 ps. Production trajectories were conducted the same as equilibration 2, however, all atoms were unrestrained during production. Production trajectories were propagated in pentaplicate from the ending of the second equilibrium and each simulation ran for 500 ns. Simulations were analyzed using the `cpptraj` module [76] to extract relevant information, including PDB files for visualization. Plots were generated using custom python scripts invoking `matplotlib` and `seaborn`. PDB files were visualized,

and figures were created, using open-source PyMOL.

4.5.18 Statistical analysis and reproducibility

No statistical methods were used to pre-determine the sample size. No sample was excluded from the data analysis. Results were reproduced using different batches of pure proteins on different days. Coomassie-stained SDS PAGE gels were done at least twice for each experiment. Microscopic fluorescence and luminescence images were repeated in biological triplicate with similar results. Unless otherwise indicated, data are shown as mean \pm s.d., and error bars in figures represent s.d. of technical triplicate. Data were analyzed and plotted using GraphPad Prism 8, seaborn, and matplotlib.

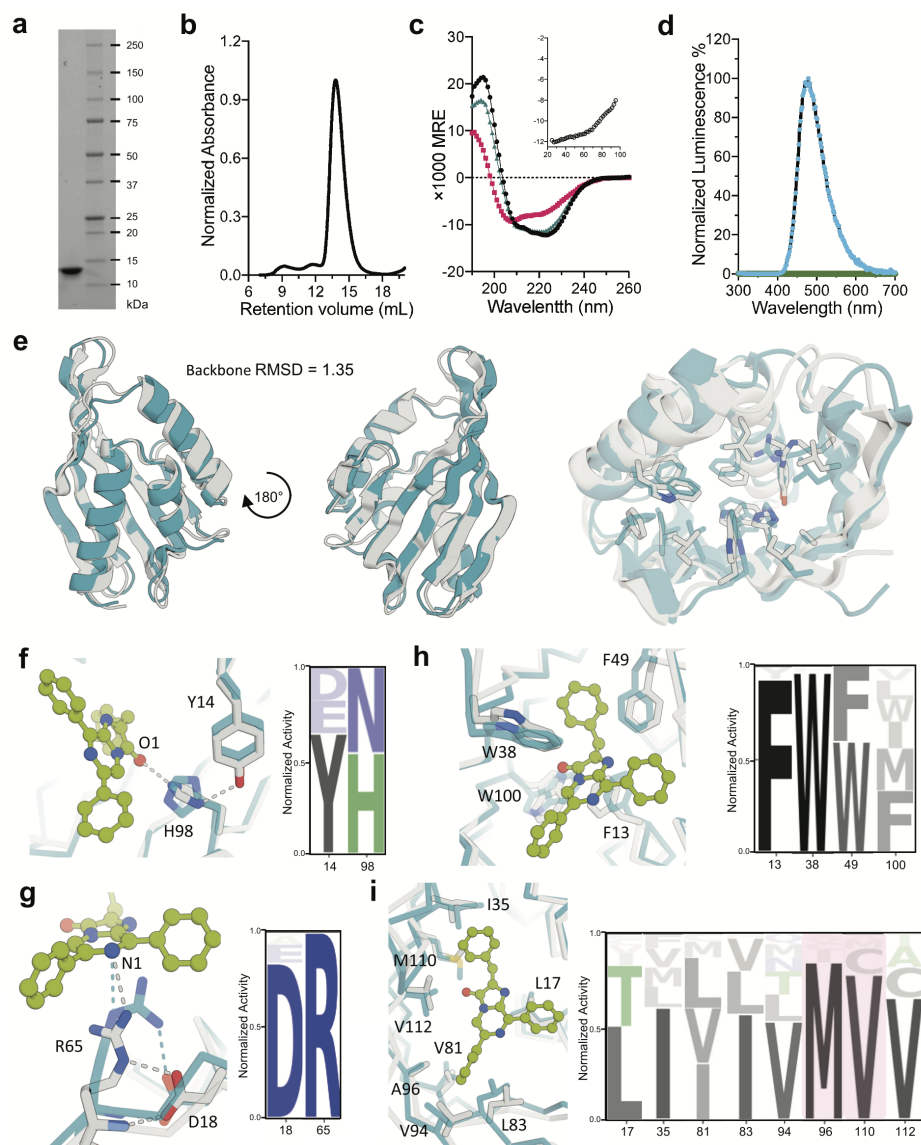
Figure 4.1: Generation of idealized scaffolds and computational design of de novo luciferases



(a) Family-wide hallucination. Sequences encoding proteins with the desired topology are optimized by Monte Carlo sampling with a multicomponent loss function. Structurally conserved regions (peach) are evaluated based on consistency with input residue-residue distance and orientation distributions obtained from 85 experimental structures of NTF2-like proteins, while variable non-ideal regions (teal) are evaluated based on the confidence of predicted inter-residue geometries calculated as the KL-divergence between network predictions and the background distribution. The sequence-space MCMC sampling incorporates both sequence changes and insertions/deletions (see Methods) to guide the hallucinated sequence toward encoding structures with the desired folds. Hydrogen-bonding networks are incorporated into the designed structures to increase structural specificity. (b-d) The design of luciferase active sites. (b) Generation of DTZ conformers. (c) Generation of a rotamer interaction field (RIF) to stabilize the anionic DTZ and form hydrophobic packing interactions around the DTZ conformers. (d) Docking of the RIF into the hallucinated scaffolds, and optimization of substrate-scaffold interactions using position-specific score matrices (PSSM)-biased sequence design. (e) Selection of the NTF2 topology. The RIF was docked into 4000 native small molecule binding proteins, excluding proteins that bind the luciferin sub-

strate using more than 5 loop residues. Most of the top hits were from the NTF2-like protein superfamily. Using the family-wide hallucination scaffold generation protocol, we generated 1615 scaffolds and found that these yielded better predicted RIF binding energies than the native proteins. **(f)** Scaffolds generated with family-wide hallucination sample more within the space of the native structures than previous Rosetta blueprint generated scaffolds, and **(g)** have the stronger sequence to structure relationships (more confident Alphafold2 structure predictions) than native or blue-print de novo NTF2 scaffolds.

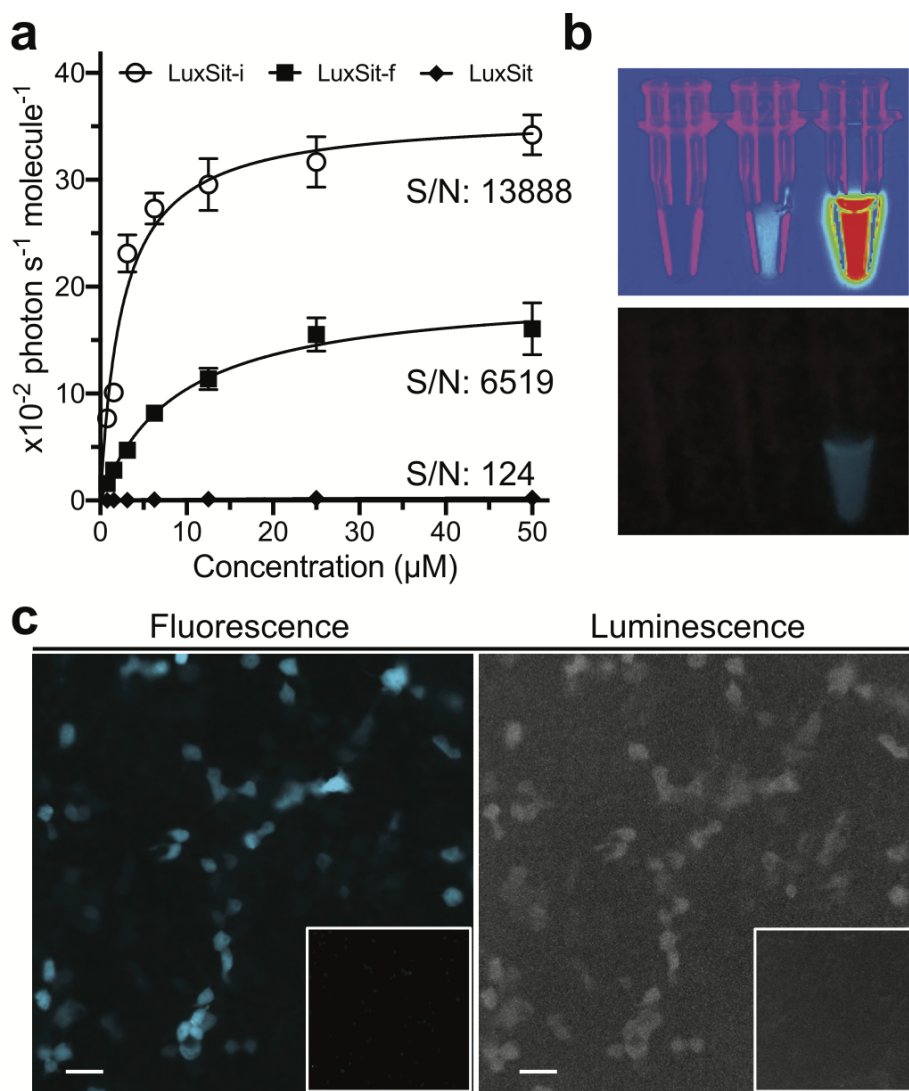
Figure 4.2: Biophysical characterization of LuxSit



(a) Coomassie-stained SDS-PAGE of purified recombinant LuxSit from *E. coli*. **(b)** Size-exclusion chromatography of purified LuxSit suggested monodispersed and monomeric properties. **(c)** Far-ultraviolet CD spectra at 25°C (black line), 95°C (red line), and cooled back to 25°C (green line). Inset: CD melting curve of LuxSit at 220 nm. **(d)** Luminescence emission spectra of DTZ in the presence (blue) and absence (green) of LuxSit. **(e)** Structural alignment of the design model (blue) and AlphaFold2 predicted model (grey), which are in close agreement at both backbone (left) and sidechain level (right). **(f-i)** Site saturation mutagenesis of substrate interacting residues. Zoomed-in views (left) of designed (blue) and AlphaFold2 models (grey) at sidechain level illustrated the designed enzyme-substrate interactions of **(f)**, Tyr14-His98 core HBNets, **(g)** Asp18-Arg65 dyad, **(h)** π -stacking, and **(i)** hydrophobic packing residues. Sequence profiles (right) are scaled by

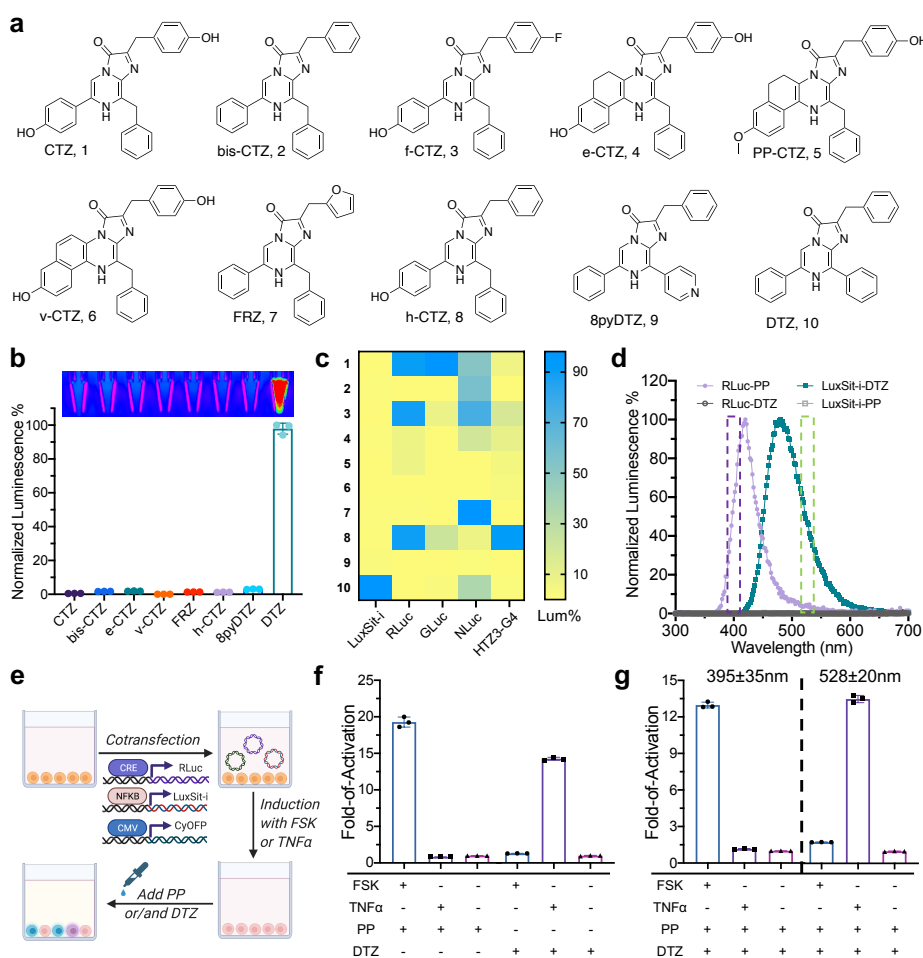
the activities of different sequence variants: (activity for the indicated amino acid) / (the sum of activities over all tested amino acids at the indicated position). A96M and M110V substitutions with increased activity are highlighted in pink.

Figure 4.3: Characterization of de novo luciferase activity in vitro and in human cells



(a) Substrate concentration dependence of LuxSit, LuxSit-f, and LuxSit-i activity. Numbers indicate the signal-to-background (S/N) ratio at V_{max}. Data are presented as mean ± s.d. (n = 3). **(b)** Luminescence images acquired by a BioRad Imager (top) or an Apple iPhone 8 (bottom). Tubes from left to right: DTZ only, DTZ plus 100 nM purified LuxSit, and DTZ plus 100 nM purified LuxSit-i, showing high efficiency of photon production. **(c)** Fluorescence and luminescence microscopic images of live HEK293T cells transiently expressing LuxSit-i-mTagBFP2; LuxSit-i activity can be detected at single-cell resolution. Left: fluorescence channel representing mTagBFP2 signal. Right: total luminescence photons were collected during a course of 10 s exposure without excitation light, immediately after adding 25 μM DTZ. Inserts: negative control, untransfected cells with DTZ. Scale bar: 20 μm. 40X.

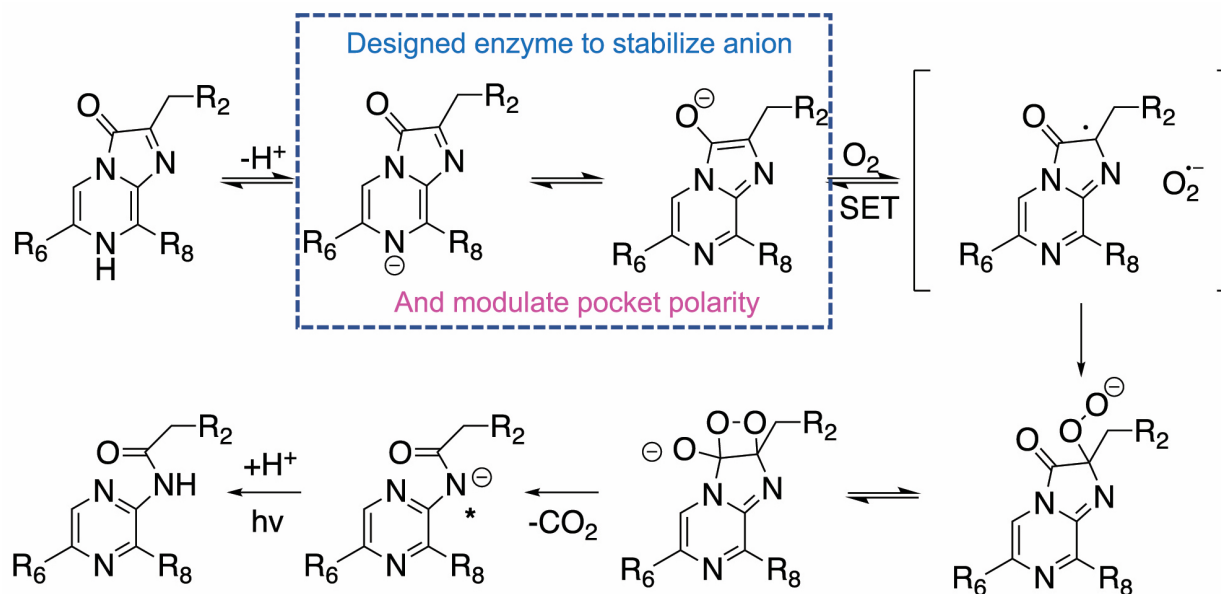
Figure 4.4: High substrate specificity of de novo luciferases allows multiplexed bioassay



(a) Chemical structures of Coelenterazine substrate analogs. **(b)** Normalized activity of LuxSit-i on selected luciferin substrates. Luminescence image (top) and signal quantification (bottom) of the indicated substrate in the presence of 100 nM LuxSit-i. LuxSit-i has high specificity for the design target substrate, DTZ. **(c)** Heatmap visualization of the substrate specificity of LuxSit-i, Renilla luciferase (RLuc), Gaussia luciferase (GLuc), engineered NLuc from Oplophorus luciferase, and the de novo luciferase (HTZ3-G4) designed for h-CTZ. The heatmap shows luminescence for each enzyme on each substrate; values are normalized on a per-enzyme basis to the highest signal for that enzyme over all substrates. **(d)** Luminescence emission spectrum of LuxSit-i/DTZ (green) and RLuc/PP-CTZ (purple) can be spectrally resolved by 528/20 and 390/35 filters (shown in dashed bars) and only recognize the cognate substrate. **(e)** Schematic of the multiplex luciferase assay. HEK293T cells transiently transfected with CRE-RLuc, NFkB-LuxSit-i, and CMV-CyOFP plasmids were treated with either Forskolin or human tumor necrosis factor alpha (TNF α) to induce the expression of labeled luciferases. **(f-g)** Luminescence signals from cells can be measured under either substrate-resolved or spectrally resolved methods by a plate reader. **(f)** For the substrate-resolved method, luminescence intensity was recorded without a filter after adding either PP-CTZ

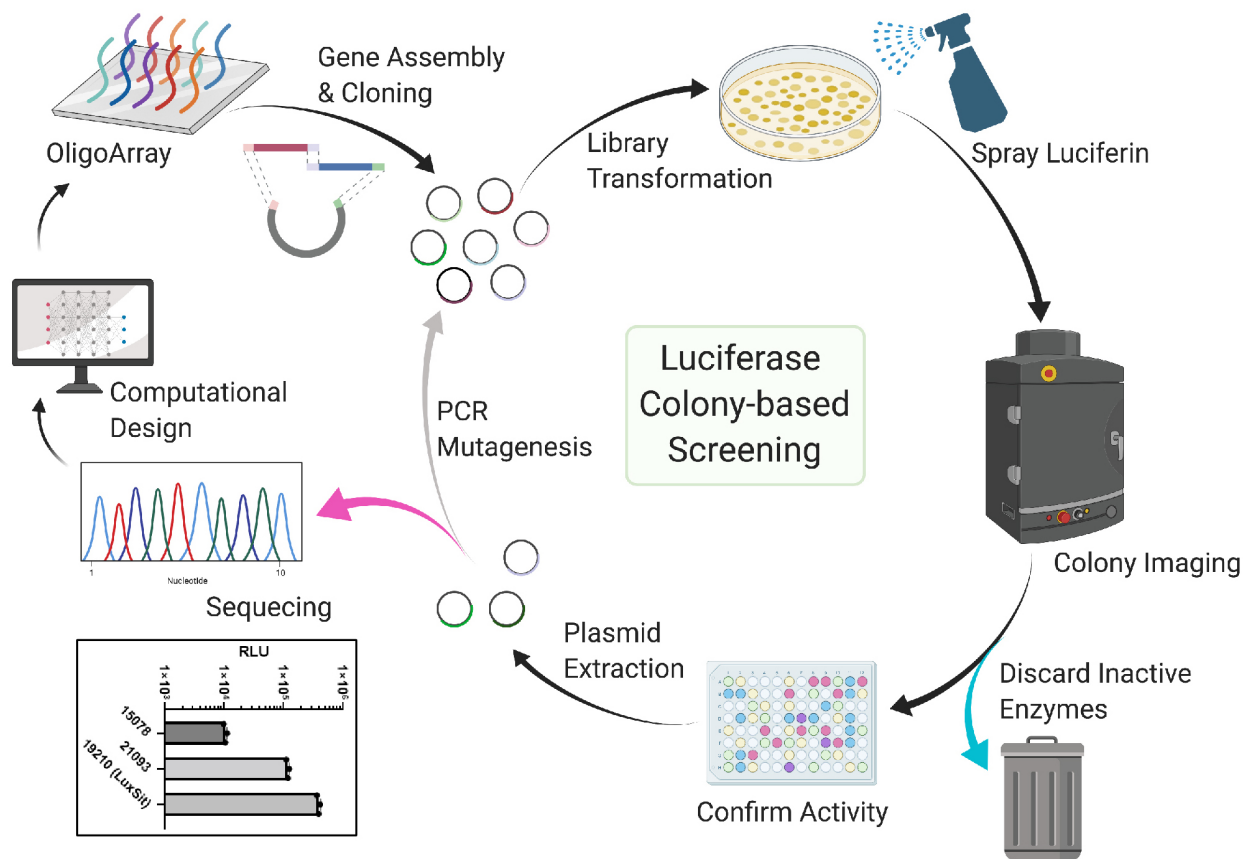
or DTZ. **(g)** For the spectrally resolved method, both PP-CTZ and DTZ were added, and the signals were acquired using 528/20 and 390/35 filters simultaneously. In **f** and **g**, the lower panel indicates the addition of Forskolin or $\text{TNF}\alpha$. Luminescence signals were acquired from the lysate of 15,000 cells in CelLytic M reagent while CyOFP fluorescence signal was used to normalize cell numbers and transfection efficiencies. All data were normalized to the corresponding non-stimulated control. Data are presented as mean \pm s.d. (n = 3, biological triplicates).

Figure E4.1: Proposed catalytic mechanism of coelenterazine-utilizing luciferases



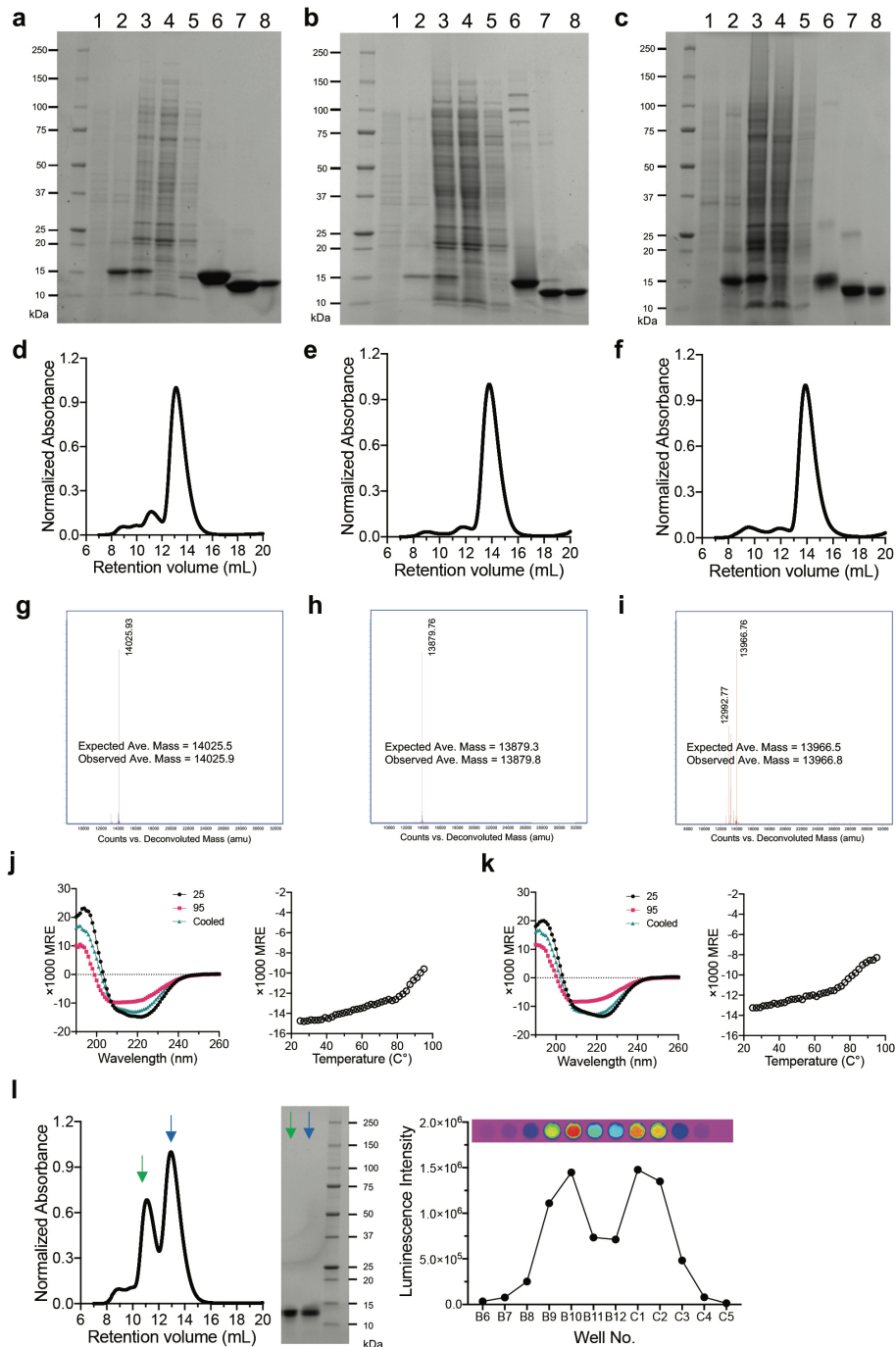
Density-functional theory (DFT) calculation suggested that the formation of an anionic state is the essential electron source for the activation of triplet oxygen (3O_2). Supported by both theoretical [25, 26] and experimental evidence [27?], the next oxygenation process is likely through a single electron transfer (SET) mechanism in which the surrounding reaction field could highly influence the change of Gibbs free energy (δG_{SET}). Finally, the thermolysis of a dioxetane light emitter intermediate can produce photons via the mechanism of gradually reversible charge-transfer-induced luminescence (GRCTIL), which is generally exergonic. As all the historical pieces of evidence are based on calculations in the virtual solvents or chemiluminescence in ideal organic solvents, the detailed mechanism of a luciferase-catalyzed luminescence reaction has remained unclear. We proposed that the key step of the enzyme is to promote the formation of an anionic state and create a suitable environment to facilitate efficient SET. Hence, the goal of this study is to design an enzyme reaction field surrounding the substrate to stabilize the anionic substrate state and alter the local proton activity, solvent polarity, and hydrophobicity for the efficient activation of 3O_2 .

Figure E4.2: Schematic representative of colony-based luciferase screening



Computationally designed DNA sequences were purchased in an oligo array, where the fragments were amplified by PCR, assembled, and ligated into a pBAD bacterial expression vector. The plasmid library was used to transform DH10B cells. Each colony grown on the LB agar plate represented one luciferase design. The plates were sprayed with DTZ solution and imaged to identify active colonies using a ChemiDoc imager. All active colonies were inoculated in 96-well plates, expressed, and purified to confirm individual luciferase activity. Selected plasmids can then be sequenced to point out active design models that provide insights into the design principle and enzyme functions or can be subjected to random mutagenesis for further evolution. Insert: three luciferases were identified from this screening. We refer to the most active and DTZ-specific luciferase as “LuxSit”.

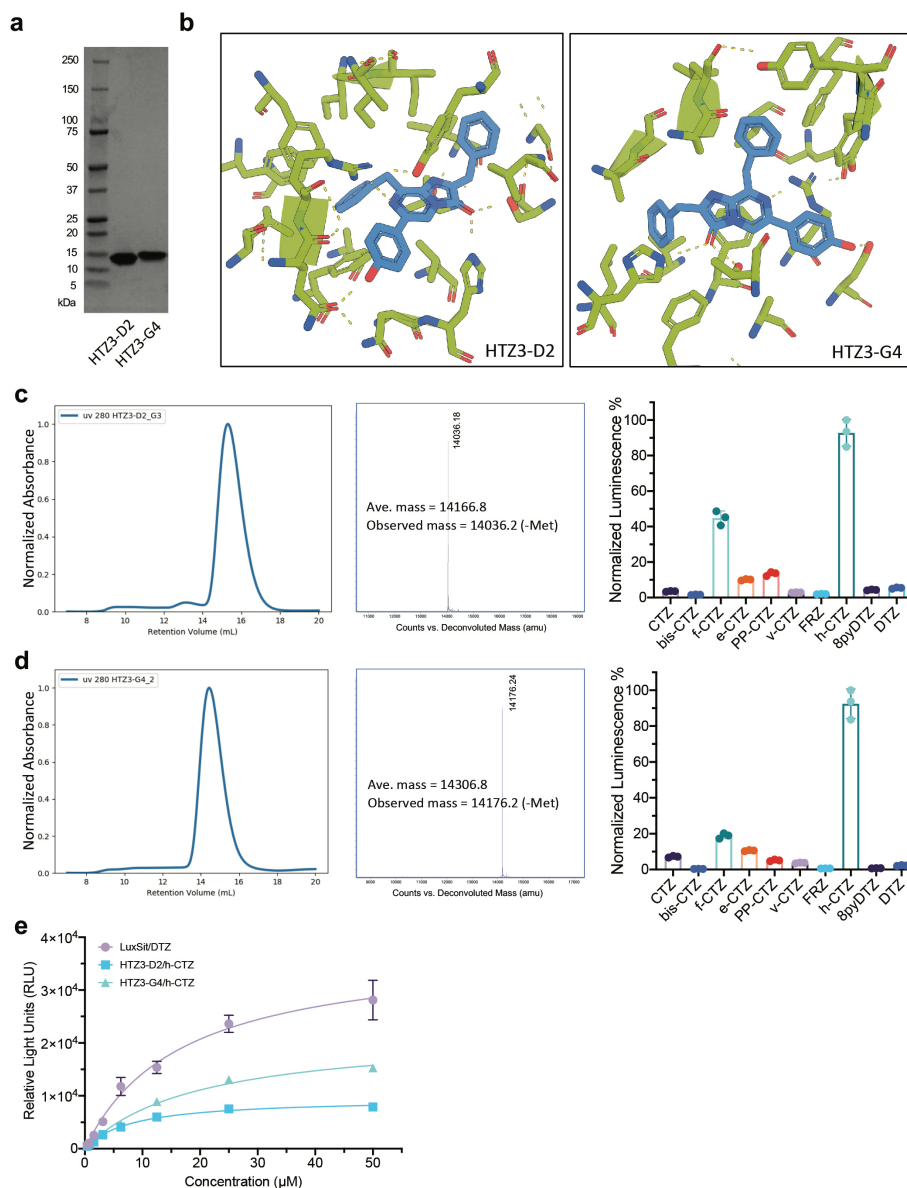
Figure E4.3: Expression, purification, and structural characterization of LuxSit variants



(a-c) The recombinant expression of (a) LuxSit, (b) LuxSit-i, and (c) LuxSit-f in *E. coli*. Annotations for each lane are the following – 1: Pre-IPTG; 2: Post-IPTG; 3: Soluble lysate; 4: Flow-through; 5: Wash; 6: Elution; 7: Post-TEV cleavage; 8: Post-SEC. (d-f) Size-exclusion chromatography of the purified (d) LuxSit; (e) LuxSit-i; and (f) LuxSit-f monomer. (g-i) Decon-

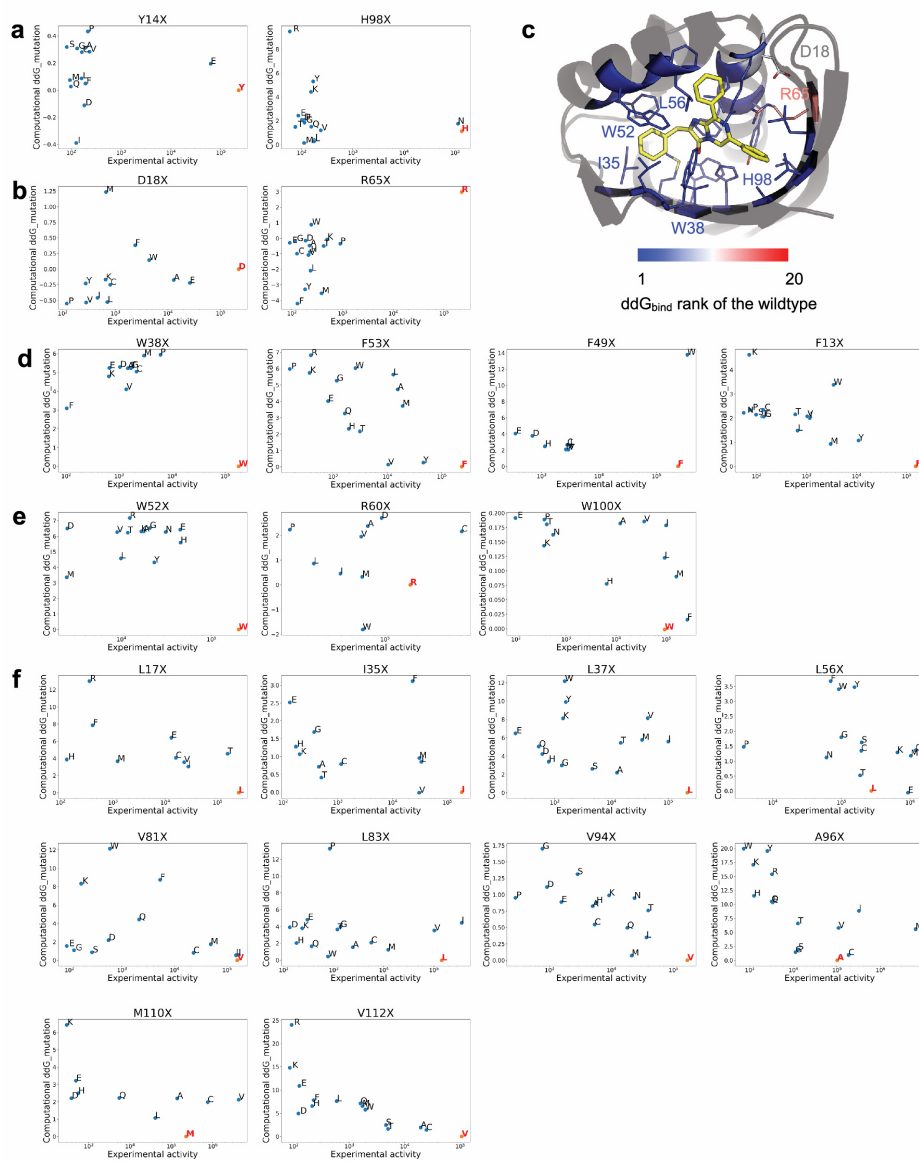
voluted mass spectrum of **(g)** LuxSit, **(h)** LuxSit-i, and **(i)** LuxSit-f. **(j-k)** Far-ultraviolet circular dichroism (CD) spectra (Left panel) of **(j)** LuxSit-i; and **(k)** LuxSit-f at 25°C (black line), 95°C (red line) and cooled back to 25°C (green line). CD melting curve at 220 nm (Right panel). **(l)** Dimeric SEC peak was observed when LuxSit-i was concentrated to high concentration (50 μ M) in Tris pH 8.0 buffer. Both dimeric and monomeric SEC fractions showed the expected size on SDS PAGE and both peaks were catalytically active to emit luminescence in the presence of 25 μ M DTZ.

Figure E4.4: Expression, purification, and activity measurement of selected de novo designed luciferases for h-CTZ



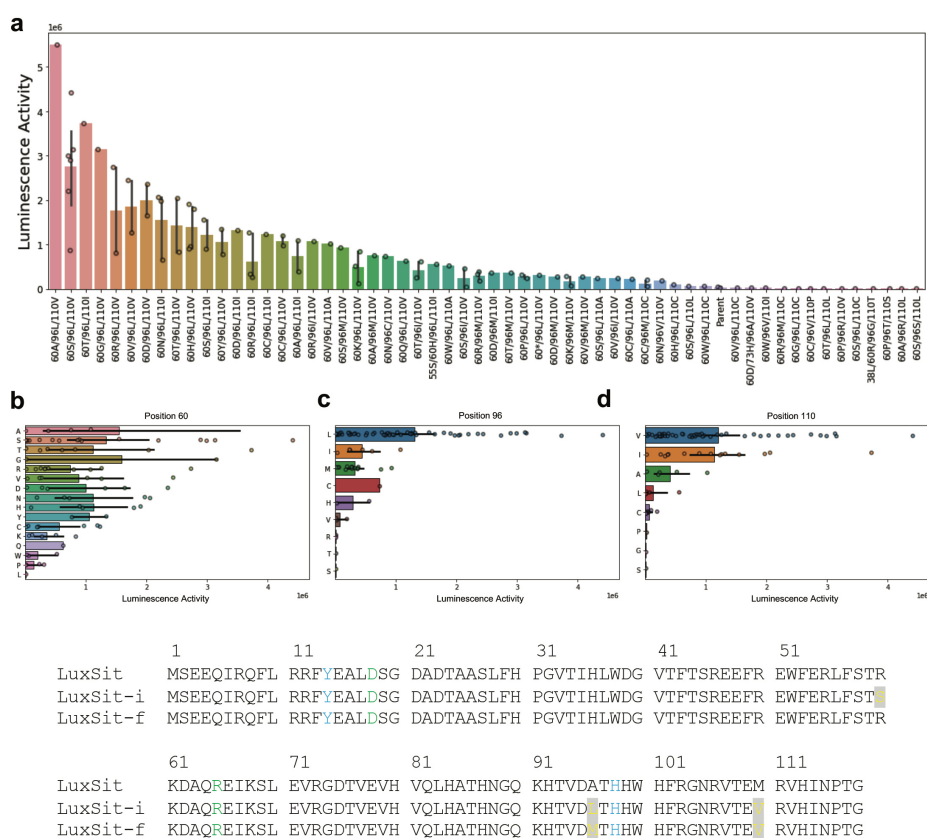
(a) Coomassie-stained SDS PAGE of HTZ3-D2 and HTZ3-G4 purified from recombinant expression in *E. coli*. **(b)** Zoomed-in views of HTZ3-D2 (left panel) and HTZ3-G4 (right panel) illustrated the sidechain preorganization of luciferase-h-CTZ interactions. **(c-d)** Size-exclusion chromatography (left), deconvoluted mass spectrum (middle), and the normalized luciferase activities on selected compounds (right) of **(c)** HTZ3-D2 and **(d)** HTZ3-G4, which suggested high specificity for the design target substrate, h-CTZ. **(e)** Substrate concentration dependence of LuxSit (w/ DTZ), HTZ3-D2 (w/ h-CTZ), and HTZ3-G4 (w/ h-CTZ) activity in PBS. All data points were fitted to Michaelis-Menten equation. HTZ3-D2 and HTZ3-G4 showed K_M values of 7.9 and 19.5 μM with 25% and 58% I_{max} of LuxSit, respectively. Data are presented as mean \pm s.d. ($n = 3$).

Figure E4.5: Predicted changes in substrate binding free energy from binding site



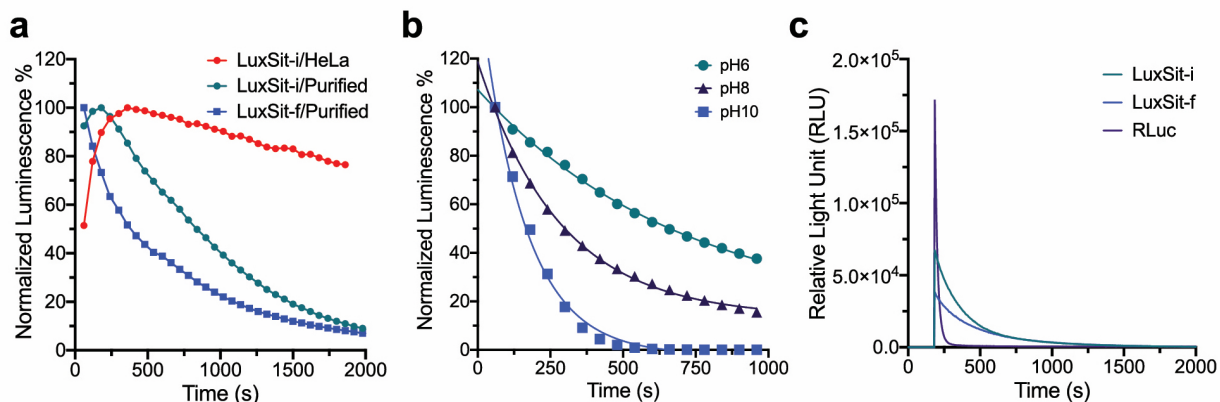
The calculated ddG_{bind} of each mutation was plotted as a function of the relative average experimental luciferase activity. The ddG_{bind} of hypothetical catalytic residues – (a) Tyr14-His98 and (b) Asp18-Arg65 dyads were generally not the lowest, which suggested that these designed catalytic residues are not favorable for substrate binding. Red dots represent the wild-type (LuxSit) amino acids. The rank of wild-type ddG_{bind} for each position screened for activity is shown with a heat map in (c) (d-f) The wild-type ddG_{bind} of the residues designed for π - π stacking or hydrophobic interactions were the lowest compared to the mutation ddG_{bind} values. This shows that the sequence is near-optimal for substrate binding and the design model is reliable.

Figure E4.6: Screening of a randomized NNK library at 60, 96, and 110 positions and sequence alignment between LuxSit and its variants



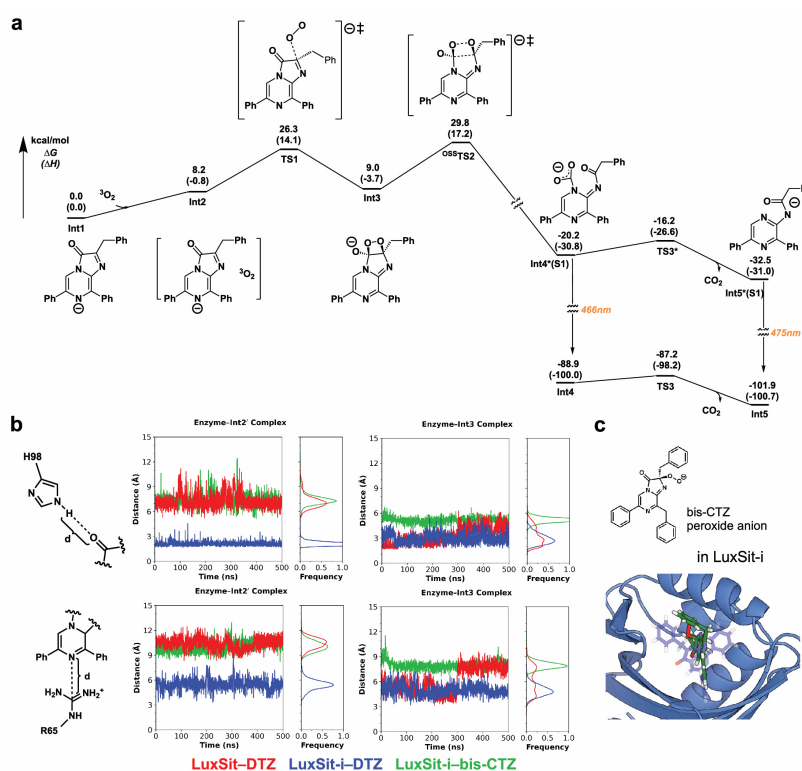
We generated a fully randomized library at 60, 96, and 110 positions to exhaustively screen all possible combinations. After the colony-based screening, we identified many colonies with strong luciferase activities with DTZ. Each colony was expressed individually in each well of 96-well plates (1 mL culture) and purified accordingly (see Methods). **(a)** Individual luminescence activity of each selected mutant was plotted and compared to the parent, LuxSit. Luminescence activities were measured in the presence of 25 μ M DTZ. Luminescence activity (RLU) was shown as the integrated signal over the first 15 min. Statistical analysis of the amino acid frequency versus the luciferase activity at residue **(b)** 60, **(c)** 96, and **(d)** 110. Data are presented as mean \pm s.d. (n varies across each bar as the mutants were selected from a randomized library). Among all selected mutants, Arg60 is confirmed to be mutable as Arg60 may be structurally less well-defined as it emanates from a loop and has no hydrogen-bonding partner. Ala96 prefers larger sidechain (Leu, Ile, Met, and Cys), and Met110 favors hydrophobic residues (Val, Ile, and Ala). A newly discovered variant (R60S/A96L/M110V) with more than 100-fold higher photon flux over LuxSit was assigned LuxSit-i for its high brightness. In the sequence alignment, mutations are highlighted in yellow fonts and gray backgrounds. The conserved catalytic dyads of Asp18-Arg65 and Tyr14-His98 are in green and blue fonts.

Figure E4.7: Additional characterization of LuxSit variants



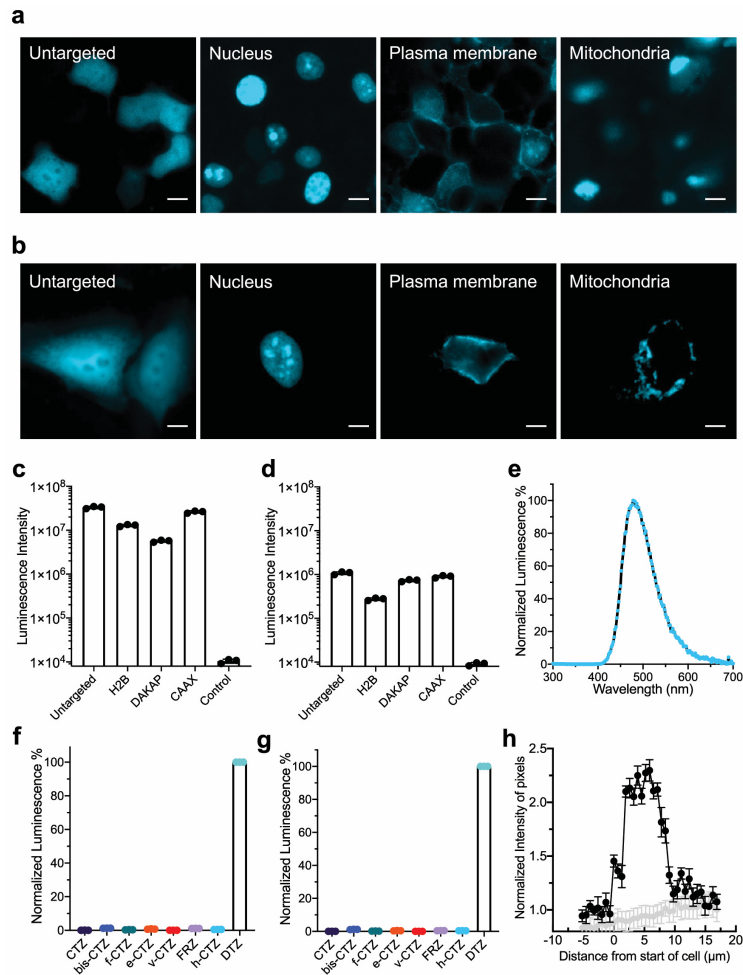
(a) Normalized emission kinetics of 15,000 intact HeLa cells expressing LuxSit-i (red), 100 nM purified LuxSit-i (green), or 100 nM purified LuxSit-f (blue) in the presence of 50 μ M DTZ. The more extended emission kinetics in HeLa cells is likely due to the diffusion rate of DTZ across cell membranes. (b) Normalized luminescence decay curves of LuxSit-i in various pH buffers revealed a pH-dependent catalytic mechanism. (c) Luminescent quantum yield was estimated from the integrated luminescence signal until completely converting 125 pmol substrates to photons in the presence of 50 nM corresponding luciferase (see Methods). Data are presented as mean ($n = 3$).

Figure E4.8: Free energy profile of DTZ chemiluminescence and molecular dynamics simulations of proposed protein-intermediate complexes



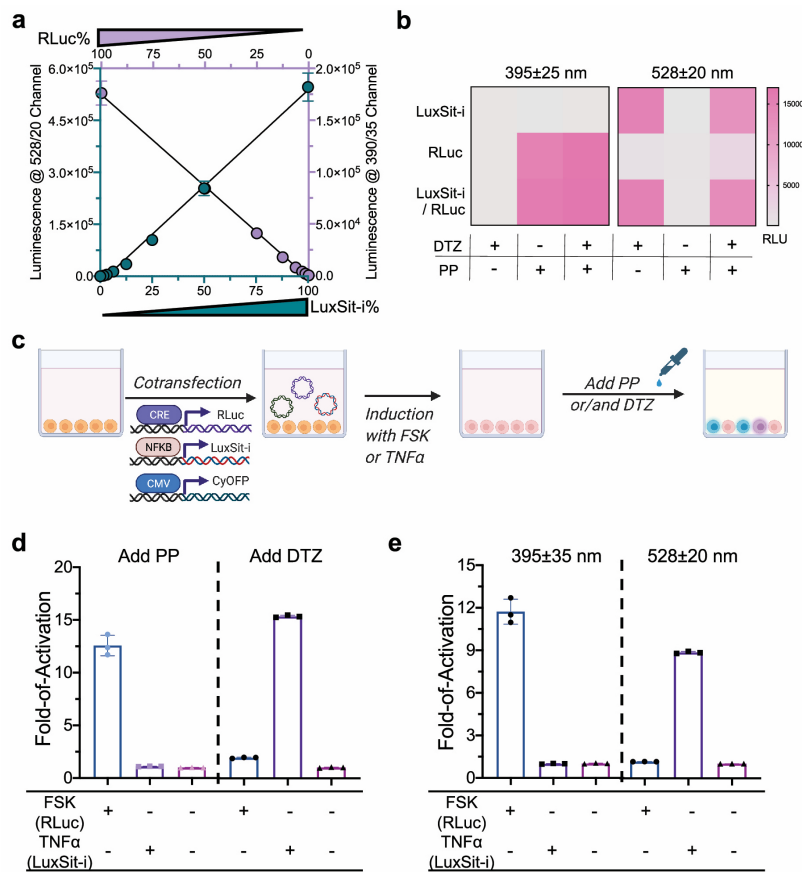
(a) The free energy profile calculated by density functional theory (DFT) shows triplet oxygen can react directly with the anionic species of DTZ (Int1) through the reactant complex Int2 and TS1. The dioxetane intermediate Int3 then cleaves in an open shell singlet transition state OSSTS2 to form excited intermediate Int4*, which rapidly converts extrudes CO₂ and forms the emissive product Int5. Note: either Int4* or Int5* emit in the observed region, but the lifetime of Int4* is very short and likely completely converts to Int5* before emission. (b) Int2 and Int3 were docked into both LuxSit and LuxSit-i and the bindings were evaluated by molecular dynamics (MD). The distances between His98 to O1 (top row) and Arg65 to N1 (bottom row) of the substrate were plotted throughout 500 ns MD simulations. LuxSit-i (blue trace) binds Int2' (middle) considerably better than LuxSit does (red trace), suggesting that the mutations of LuxSit-i provide a binding pocket more complimentary to TS1. This binding orientation brings N1 of the substrate much closer to Arg65, providing better charge stabilization for the high energy transition state. (c) Docking of the peroxide anion form of bis-CTZ into the pocket of LuxSit-i; blue overlay represents DTZ in the original design model. During MD simulation, the added benzylic carbon of bis-CTZ (green trace) disrupts the shape complementarity between LuxSit-i and the transition states (TS1 and TS2), reducing the charge stabilization by Arg65. This charge stabilization is necessary for the reaction to proceed, explaining the high substrate specificity of LuxSit-i for DTZ over bis-CTZ.

Figure E4.9: Expression, localization, and luminescence activity of LuxSit-i in live HEK293T and HeLa cells



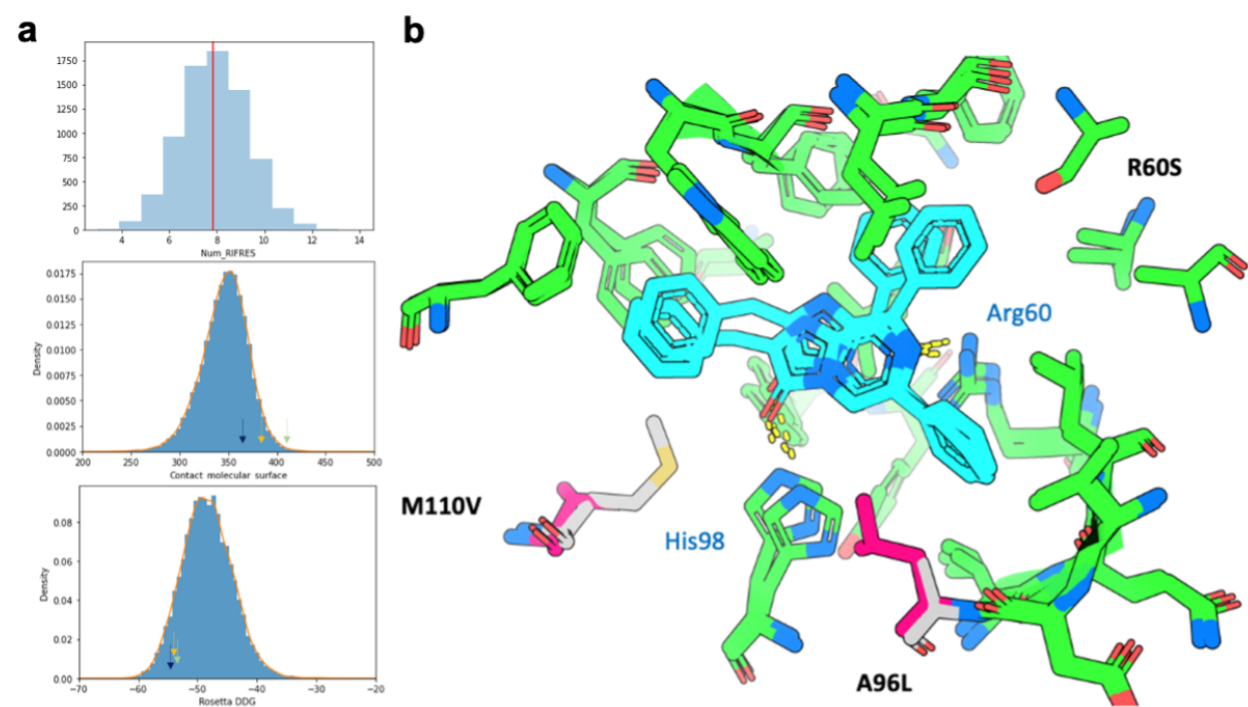
(a)-(b) Fluorescence imaging of live (a) HEK293T and b, HeLa cells expressing LuxSit-i-mTagBFP2, which is untargeted or localized to the nucleus (Histone2B), plasma membrane (KRasCAAX), or mitochondria (DAKAP) cellular compartments. Scale bar: 10 μ m. (c-d) Luminescence signals were measured with 15,000 intact (c) HEK293T or (d) HeLa cells in the presence of 25 μ M DTZ in DPBS. Transfection efficiencies range from 60-70% for HEK293T cells and 5-10% for HeLa cells. (e) Luminescence emission spectra acquired from LuxSit-i expressing HEK293T cells is consistent with the emission spectra of recombinant LuxSit-i purified from *E. coli*. (f-g) Luminescence signals were measured with 15,000 (f) intact LuxSit-i expressing HEK293T cells or (g) cell lysate in the presence of 25 μ M indicated substrate in DPBS. Luminescence intensities were normalized to DTZ signal, showing high DTZ specificity over other substrates in cell-based assays. Data were shown as total luminescence signal over the first 20 min \pm s.d. (n = 3, biological triplicates). (h) Normalized luminescence intensity profile of lines traversing across different cells (n=10) of main Figure 4.3c luminescence image; grey lines represent untransfected cells. Error bars represent \pm SEM.

Figure E4.10: Substrate specificity of LuxSit-i and spectrally resolved luciferase-luciferin pairs allow multiplexed bioassay



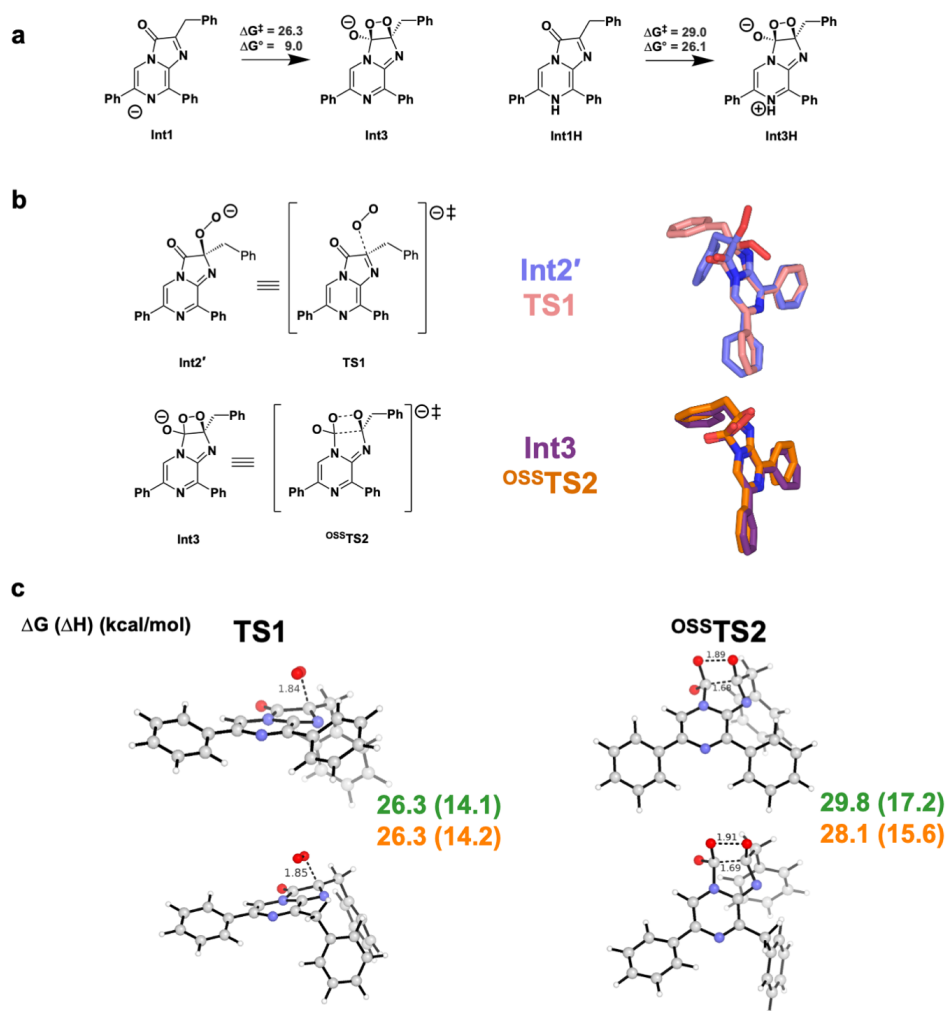
(a) The orthogonality relationship between LuxSit-i-DTZ and RLuc-PP-CTZ (Prolume Purple, methoxy e-Coelenterazine) luminescent pairs. Indicated percentages of each luciferase were mixed at different ratios totaling 100%. After the addition of both 25 μ M DTZ and PP-CTZ substrates, filtered light from 528/20 and 390/35 channels were measured simultaneously. (b) Heatmap shows the luminescence signal for individual luciferase (100 nM) or 1:1 mixture in the presence of the cognate or non-cognate (DTZ or PP-CTZ or both) substrates. Response signals were acquired by a Neo2 plate reader with 528/20 and 390/35 nm filters simultaneously. (c) Multiplex luciferase assay in live HEK293T after co-transfection of CRE-RLuc, NF κ B-LuxSit-i, and CMV-CyOFP plasmids and stimulation by Forskolin (FSK) or human tumor necrosis factor alpha (TNF α). (d,e) 15,000 intact cells were assayed (see Methods) by either (d) substrate-resolved or (e) spectrally resolved modes after adding DTZ, PP-CTZ, or both DTZ and PP-CTZ in DPBS without cell lysis. Area scanning of the CyOFP fluorescence signal was used to estimate cell numbers and transfection efficiency. The reported unit was RLU/a.u.; relative light units/fluorescence intensity measurements at Ex./Em.=480/580 nm. All data were normalized to the corresponding non-stimulated control. Data are presented as mean \pm s.d. (n = 3, biological triplicates).

Figure S4.1: Additional computational analysis of designed luciferases



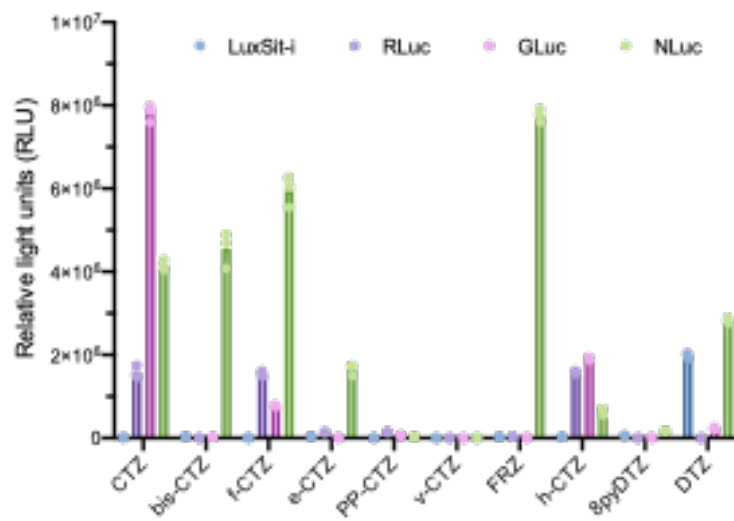
(a) Histogram plots of 7648 luciferase designs show the number of sidechains placed on each scaffold by RIFdock (top, the red line indicates the average) and in silico scores, such as contact molecular surface (middle) and Rosetta ddG (bottom). Three active DTZ designs we experimentally identified (blue, green, and yellow arrows indicate 15078, 21093, and LuxSit designs, see Figure E4.2) were located on the upper end of both distributions, suggesting that having substantial packing and contact with the ligand is an important factor for active luciferases. (b) Superimposed protein sidechain and DTZ interactions of LuxSit and Luxsit-i. Both models were generated by Rosetta FastRelax protocol. Mutations (R60S, A96L, and M110V) of LuxSit-i were colored in magenta. Full atom RMSD = 1.2 Å. For the ligand interactions of LuxSit, ten sidechains were initially placed from RIFdock, and five of them were mutated during RosettaDesign. Both Tyr-His and Asp-Arg dyads were from the pre-defined hydrogen bond networks.

Figure S4.2: Additional QM and MM computational models used in this study



(a) Comparison of energies for the first proposed step indicates that deprotonation of N1 is critical for the reaction to proceed. (b) Overlay of Int2' and TS1 (top) and overlay of Int3 and ^{OSST2} highlighting the structural similarities between transition states and their analogs. While Int2' does not lie on the intrinsic reaction path connecting TS1 and Int3, its structural similarity to TS1 makes it a useful transition state analog for ground state MD simulations. (c) Comparison of TS1 (left) and ^{OSST2} (right) for DTZ (top) and bis-CTZ (bottom) pathways. Computed transition state energies are nearly identical, suggesting that the observed selectivity is entirely due to the protein.

Figure S4.3: The raw light outputs plot of Figure 4.4c



Indicated luciferin analogs (25 μ M) were mixed with either LuxSit-i, RLuc, GLuc, or NLuc (100 nM). The total luminescence signal over the first 10 min was plotted (n=3). We normalized the data by setting the highest emission substrate at 100% in each group to create Figure 4.4c.

4.6 Computational Models

Table S4.1: Computed Energies of Computational Models (Hartree)

Structure	Electronic Energy	Zero Point Energy	Enthalpy	Free Energy
Reactant Complex	-996.17861	0.221778	-995.94114	-995.997464
NAC5	-996.17076	0.221113	-995.93611	-995.9937
TS5	-996.14381	0.216471	-995.91225	-995.967078
NAC6	-996.17077	0.22101	-995.93611	-995.99385
TS6	-996.143661	0.216713	-995.91189	-995.968527
NAC7	-996.172	0.221115	-995.93751	-995.99517
TS7	-996.142862	0.21693	-995.91101	-995.965437

Cartesian Coordinates

1	C 0.143600 3.381629 -0.672305
C -1.246644 -1.594445 -0.567477	C 0.184212 4.759387 -0.473300
C -2.280262 -0.781826 -0.211328	C -0.757273 5.382431 0.341829
N -2.122882 0.579101 -0.054370	C -1.745849 4.616543 0.958364
C -0.941878 1.131896 -0.260803	C -1.790566 3.241975 0.757583
C 0.183991 0.357002 -0.640116	H -2.560298 2.641400 1.230539
N 1.457040 0.655089 -0.876585	H -2.483280 5.091824 1.598819
C 2.066919 -0.520477 -1.188375	H -0.721491 6.456797 0.497074
C 1.185188 -1.631482 -1.140727	H 0.955864 5.348133 -0.961016
O 1.325856 -2.871140 -1.326815	H 0.885331 2.898366 -1.296908
N -0.021815 -1.024792 -0.792547	C -3.637491 -1.322901 0.053612
C 3.533766 -0.612930 -1.486262	C -4.507006 -0.638525 0.911663
C 4.378398 -0.878566 -0.252662	C -5.776457 -1.136928 1.186744
C 4.185323 -2.053761 0.484072	C -6.204863 -2.329412 0.606829
C 4.941342 -2.309804 1.623110	C -5.352607 -3.013932 -0.257069
C 5.903595 -1.392854 2.049017	C -4.084719 -2.512643 -0.535688
C 6.100459 -0.220534 1.325358	H -3.447066 -3.044192 -1.236363
C 5.341532 0.032288 0.182020	H -5.679677 -3.936574 -0.727730
H 5.500781 0.951001 -0.377982	H -7.196162 -2.718090 0.820114
H 6.845119 0.501370 1.649192	H -6.433221 -0.592274 1.859204
H 6.493487 -1.592554 2.938940	H -4.172497 0.290398 1.361721
H 4.780820 -3.227321 2.182805	H -1.293646 -2.671126 -0.672308
H 3.426593 -2.758749 0.151339	
H 3.869119 0.313021 -1.966342	2
H 3.692856 -1.427344 -2.204631	C -1.198752 -1.621041 -0.544589
C -0.845936 2.604759 -0.057317	C -2.234362 -0.802332 -0.210112

N -2.072450 0.559212 -0.054816
C -0.886706 1.103336 -0.247013
C 0.241670 0.320035 -0.602661
N 1.520616 0.616171 -0.818495
C 2.129082 -0.556532 -1.122996
C 1.242379 -1.670323 -1.078283
O 1.387335 -2.907816 -1.249327
N 0.033743 -1.058924 -0.744130
C 3.597022 -0.655862 -1.409283
C 4.428561 -0.901521 -0.163015
C 4.236780 -2.070408 0.583704
C 4.979452 -2.306621 1.735769
C 5.926837 -1.375682 2.164311
C 6.122344 -0.209425 1.430449
C 5.376637 0.023673 0.274433
H 5.534354 0.937752 -0.293416
H 6.855657 0.523081 1.756252
H 6.506439 -1.559814 3.064271
H 4.820026 -3.219447 2.303304
H 3.489673 -2.786736 0.249158
H 3.936365 0.261891 -1.901659
H 3.758443 -1.482399 -2.112663
C -0.783117 2.576889 -0.064595
C 0.174635 3.343318 -0.739321
C 0.222900 4.723895 -0.564138
C -0.679114 5.358231 0.286412
C -1.636643 4.602147 0.961329
C -1.689729 3.224503 0.783763
H -2.435903 2.630649 1.301424
H -2.342974 5.087434 1.628667
H -0.637198 6.434857 0.423279
H 0.968290 5.306064 -1.098269
H 0.882685 2.850083 -1.394391
C -3.599457 -1.334843 0.029459
C -4.481245 -0.643579 0.869177
C -5.758328 -1.134634 1.121139
C -6.181994 -2.326071 0.535671
C -5.317352 -3.017049 -0.310590
C -4.041684 -2.523303 -0.566030
H -3.393790 -3.059835 -1.253370
H -5.640493 -3.938846 -0.785547
H -7.179228 -2.708964 0.730925
H -6.424816 -0.584973 1.779779
H -4.150608 0.284691 1.323329
H -1.249998 -2.697568 -0.648572

O 0.549982 -0.374891 -3.734206
O -0.199959 0.538438 -3.456816

TS1

C -0.793428 -1.493122 -0.373176
C -1.821085 -0.683578 -0.010039
N -1.631736 0.682890 0.163252
C -0.475672 1.231075 -0.061664
C 0.666119 0.443270 -0.497134
N 1.952039 0.754506 -0.564888
C 2.494257 -0.343053 -1.159296
C 1.626838 -1.536119 -0.981961
O 1.801825 -2.737403 -1.151949
N 0.436692 -0.934145 -0.564463
C 3.963743 -0.466054 -1.421918
C 4.749139 -0.719485 -0.152198
C 4.740616 -1.984482 0.444727
C 5.435575 -2.215598 1.628315
C 6.150537 -1.183356 2.234378
C 6.162373 0.080207 1.648952
C 5.464524 0.308564 0.464313
H 5.475800 1.297368 0.012486
H 6.716234 0.890947 2.113650
H 6.694793 -1.363637 3.156759
H 5.420428 -3.204184 2.078437
H 4.172339 -2.784627 -0.022633
H 4.303412 0.455455 -1.903731
H 4.114475 -1.288063 -2.129904
C -0.340285 2.697476 0.095498
C 0.469293 3.433025 -0.777384
C 0.557070 4.815569 -0.645528
C -0.151003 5.471896 0.359444
C -0.960985 4.742051 1.228722
C -1.060131 3.361715 1.093733
H -1.691290 2.785444 1.763283
H -1.514727 5.249474 2.012993
H -0.074158 6.550109 0.464475
H 1.178793 5.383242 -1.331586
H 1.002264 2.907376 -1.562297
C -3.190568 -1.194277 0.232584
C -4.059400 -0.489711 1.074680
C -5.342058 -0.963941 1.329787
C -5.782764 -2.149841 0.746055
C -4.930229 -2.852986 -0.102453
C -3.649149 -2.376813 -0.362333

H -3.012097 -2.922147 -1.052840
H -5.267183 -3.770177 -0.576398
H -6.784543 -2.519005 0.943914
H -5.999352 -0.404873 1.989466
H -3.717060 0.434681 1.527915
H -0.848949 -2.568180 -0.492312
O 1.691797 -0.276226 -2.809036
O 0.644392 0.547987 -2.723222

3

C -0.692973 -1.488501 -0.897862
C -1.621249 -0.744577 -0.213042
N -1.347523 0.562596 0.078042
C -0.227287 1.141137 -0.268214
C 0.794176 0.390765 -1.000906
N 1.998818 0.755158 -1.343481
C 2.605392 -0.371690 -2.005004
C 1.642171 -1.617217 -1.959143
O 1.835322 -2.788971 -1.556661
N 0.434663 -0.899067 -1.310745
C 4.073509 -0.560694 -1.663970
C 4.351709 -0.499940 -0.179501
C 3.776441 -1.434448 0.689847
C 4.020849 -1.372297 2.058489
C 4.847381 -0.378687 2.582551
C 5.424899 0.554020 1.725416
C 5.175195 0.492243 0.355191
H 5.625750 1.227806 -0.307028
H 6.069525 1.333567 2.121729
H 5.038345 -0.332192 3.650823
H 3.564463 -2.103157 2.720621
H 3.122670 -2.199364 0.275986
H 4.647591 0.214360 -2.184907
H 4.385397 -1.529046 -2.072353
C -0.033532 2.558107 0.135172
C 0.946013 3.377971 -0.440821
C 1.071113 4.709483 -0.050071
C 0.230264 5.243345 0.921299
C -0.745765 4.433858 1.503072
C -0.876931 3.108055 1.112984
H -1.637274 2.479066 1.563044
H -1.405387 4.837737 2.265599
H 0.333924 6.280414 1.226582
H 1.833929 5.330852 -0.510358
H 1.613904 2.961904 -1.184454

C -2.910166 -1.296983 0.260959
C -3.576663 -0.696303 1.336614
C -4.780404 -1.211295 1.805538
C -5.343458 -2.338530 1.210248
C -4.693052 -2.940084 0.135136
C -3.492258 -2.421206 -0.339473
H -3.018117 -2.888106 -1.198146
H -5.127330 -3.811213 -0.346593
H -6.282887 -2.741053 1.576889
H -5.278807 -0.731912 2.643220
H -3.137701 0.179130 1.803989
H -0.779217 -2.544757 -1.129487
O 2.419304 -0.313308 -3.439322
O 1.476583 -1.434985 -3.439673

OSS TS2

C -0.743517 -1.502981 -0.758279
C -1.679100 -0.728214 -0.123280
N -1.393560 0.582700 0.147835
C -0.269341 1.141466 -0.210530
C 0.741249 0.372468 -0.944395
N 1.913622 0.746525 -1.356075
C 2.511513 -0.368023 -2.058751
C 1.616327 -1.740812 -1.700360
O 1.947081 -2.681906 -0.940676
N 0.400082 -0.940008 -1.176864
C 4.016410 -0.521822 -1.813801
C 4.423222 -0.509747 -0.360577
C 4.662326 -1.700740 0.329467
C 5.035745 -1.687699 1.671980
C 5.176925 -0.477914 2.347299
C 4.942919 0.717668 1.669322
C 4.571167 0.698810 0.328139
H 4.383442 1.632091 -0.194962
H 5.052924 1.667136 2.185900
H 5.469582 -0.465517 3.393318
H 5.217296 -2.625200 2.190281
H 4.538173 -2.645378 -0.190539
H 4.514703 0.298867 -2.344334
H 4.328257 -1.455874 -2.291781
C -0.051584 2.557501 0.186369
C 0.924744 3.368124 -0.407855
C 1.073793 4.696657 -0.015182
C 0.260203 5.236764 0.975725
C -0.713136 4.436888 1.574936

C -0.867813 3.114060 1.183380
H -1.625304 2.492206 1.648005
H -1.351567 4.845575 2.352767
H 0.382864 6.271305 1.282576
H 1.833850 5.310712 -0.489609
H 1.569513 2.947183 -1.168860
C -2.989266 -1.250667 0.324741
C -3.667203 -0.630377 1.381978
C -4.892030 -1.117581 1.825171
C -5.465012 -2.235702 1.221949
C -4.803439 -2.855982 0.164517
C -3.581522 -2.364520 -0.284835
H -3.098669 -2.842911 -1.132332
H -5.245552 -3.719401 -0.323923
H -6.421121 -2.616210 1.568408
H -5.399508 -0.623711 2.648873
H -3.220428 0.238119 1.854971
H -0.847228 -2.565789 -0.946873
O 2.165548 -0.433470 -3.359592
O 1.399727 -2.134309 -3.031020

4

C -1.621756 -1.849065 -0.646394
C -2.467111 -0.970265 -0.038579
N -1.976823 0.220148 0.433052
C -0.749542 0.568320 0.182596
C 0.142749 -0.251830 -0.655107
N 1.264315 0.239495 -1.067371
C 2.042652 -0.245296 -2.081214
C 0.589390 -2.716807 -1.332995
O 1.744411 -2.624702 -0.919758
N -0.322539 -1.530680 -0.904114
C 3.547093 -0.247589 -1.811036
C 3.966728 0.261158 -0.456434
C 3.786301 -0.529828 0.683243
C 4.137097 -0.051278 1.941892
C 4.675936 1.227933 2.082693
C 4.861913 2.021290 0.953662
C 4.510672 1.537350 -0.305638
H 4.655756 2.164055 -1.182573
H 5.282228 3.018255 1.051943
H 4.948035 1.602658 3.065208
H 3.985930 -0.675496 2.818262
H 3.334513 -1.511075 0.569922
H 4.022257 0.329871 -2.612088

H 3.857989 -1.288805 -1.946347
C -0.282120 1.860782 0.754956
C 1.019467 2.043508 1.238930
C 1.382130 3.245316 1.839149
C 0.460970 4.284402 1.954129
C -0.834326 4.112106 1.470354
C -1.203363 2.907428 0.881420
H -2.213744 2.766219 0.511072
H -1.560047 4.916034 1.551191
H 0.750669 5.223534 2.416427
H 2.394135 3.364850 2.214968
H 1.752555 1.250948 1.146229
C -3.892782 -1.273055 0.219532
C -4.560190 -0.649091 1.280741
C -5.897027 -0.929889 1.541975
C -6.593339 -1.839616 0.748526
C -5.940910 -2.460147 -0.314595
C -4.605313 -2.174664 -0.581391
H -4.122606 -2.642456 -1.434995
H -6.477092 -3.159257 -0.949570
H -7.637518 -2.057226 0.951744
H -6.395714 -0.439208 2.372880
H -4.017498 0.056507 1.901043
H -1.895122 -2.854867 -0.936702
O 1.621667 -0.549094 -3.193355
O -0.019782 -3.574004 -1.969216

TS3

C -1.675115 -1.867084 -0.690085
C -2.524601 -0.977022 -0.062905
N -2.017644 0.178441 0.424594
C -0.754861 0.489197 0.222191
C 0.116305 -0.369105 -0.563005
N 1.362216 0.006433 -0.826872
C 2.060142 -0.421071 -1.895433
C 0.699085 -3.141291 -1.415998
O 1.692213 -3.059702 -0.763851
N -0.391837 -1.575475 -0.923166
C 3.584044 -0.265203 -1.765840
C 4.068027 0.325284 -0.468886
C 4.127369 -0.459791 0.686878
C 4.523967 0.086998 1.903038
C 4.877911 1.434044 1.983122
C 4.832594 2.223193 0.836787
C 4.430909 1.669778 -0.377844

H 4.391025 2.294747 -1.267106
H 5.108571 3.272844 0.887097
H 5.189661 1.862766 2.931193
H 4.558688 -0.537924 2.791188
H 3.836769 -1.505727 0.629047
H 3.928260 0.329257 -2.620054
H 3.995939 -1.272052 -1.905522
C -0.288864 1.772043 0.823425
C 1.020270 1.966903 1.286215
C 1.378858 3.163264 1.901396
C 0.449243 4.188833 2.055150
C -0.853248 4.006812 1.593841
C -1.217698 2.808631 0.989791
H -2.234153 2.662604 0.638919
H -1.587485 4.799877 1.703179
H 0.737079 5.123845 2.527259
H 2.398431 3.289439 2.254411
H 1.761041 1.188801 1.147894
C -3.962727 -1.252972 0.159233
C -4.644323 -0.626794 1.210521
C -5.993053 -0.880810 1.434509
C -6.688085 -1.766574 0.612757
C -6.022041 -2.389537 -0.440170
C -4.673734 -2.131176 -0.668826
H -4.180009 -2.602685 -1.513829
H -6.555834 -3.071361 -1.095682
H -7.741064 -1.965324 0.788210
H -6.502927 -0.388002 2.257333
H -4.102187 0.060567 1.851677
H -1.993586 -2.855885 -1.007554
O 1.612224 -0.852707 -2.966685
O 0.007334 -3.742269 -2.178630

5

C -0.861297 -1.525589 0.043247
C -1.926050 -0.644281 0.203121
N -1.696670 0.677383 0.110472
C -0.472110 1.109536 -0.154493
C 0.627278 0.195283 -0.327153
N 1.886110 0.676211 -0.466332
C 2.805643 0.053458 -1.217284
N 0.382743 -1.129694 -0.212925
C 4.239158 0.542360 -0.935603
C 4.887579 -0.357950 0.089696
C 5.662164 -1.453276 -0.301932

C 6.217516 -2.311562 0.644935
C 6.002525 -2.088336 2.003086
C 5.226911 -1.001952 2.405609
C 4.673933 -0.147305 1.455947
H 4.062070 0.693928 1.771261
H 5.053156 -0.819380 3.462451
H 6.437127 -2.754380 2.742770
H 6.820317 -3.155277 0.320594
H 5.831069 -1.633356 -1.360689
H 4.211338 1.571845 -0.567736
H 4.806706 0.514850 -1.870937
C -0.315419 2.589124 -0.232599
C 0.531693 3.209531 -1.159219
C 0.608112 4.597589 -1.231615
C -0.150778 5.391701 -0.374340
C -0.997297 4.785384 0.551677
C -1.081711 3.398237 0.615852
H -1.747651 2.925174 1.331035
H -1.593143 5.393638 1.226279
H -0.083257 6.474451 -0.427663
H 1.264885 5.060659 -1.962662
H 1.133506 2.598083 -1.820354
C -3.308778 -1.088423 0.502165
C -4.193227 -0.226749 1.163985
C -5.490691 -0.628311 1.462200
C -5.932243 -1.901433 1.105065
C -5.064632 -2.764663 0.440114
C -3.767246 -2.361055 0.137268
H -3.116379 -3.039062 -0.407157
H -5.400702 -3.754434 0.144968
H -6.945004 -2.215868 1.338779
H -6.159050 0.053986 1.979696
H -3.848218 0.763604 1.443534
H -1.006669 -2.599096 0.155302
O 2.642924 -0.813409 -2.093467

TS3*

C -1.576489 -1.810273 -0.649267
C -2.465556 -0.895093 -0.048638
N -2.061299 0.303853 0.392872
C -0.722722 0.621990 0.215504
C 0.107875 -0.219405 -0.611364
N 1.256257 0.242368 -1.072429
C 2.008682 -0.312123 -2.069227
C 0.856356 -3.092119 -1.375595

O 1.898941 -2.867132 -0.851479
N -0.309100 -1.486672 -0.936547
C 3.519078 -0.305480 -1.830229
C 3.965881 0.239844 -0.498290
C 3.835711 -0.527896 0.663290
C 4.204172 -0.010890 1.901759
C 4.714595 1.282835 1.997975
C 4.856007 2.052519 0.845935
C 4.482704 1.532255 -0.391756
H 4.588339 2.141491 -1.286250
H 5.255419 3.060734 0.909601
H 5.001775 1.687313 2.964078
H 4.092007 -0.618770 2.795074
H 3.424156 -1.530404 0.590203
H 3.970231 0.258372 -2.654461
H 3.838116 -1.346102 -1.951771
C -0.272749 1.845496 0.838407
C 1.089304 2.186898 1.058595
C 1.437386 3.366929 1.695263
C 0.460826 4.255482 2.156542
C -0.884594 3.928380 1.975038
C -1.248101 2.751276 1.339554
H -2.294386 2.506645 1.198700
H -1.659340 4.602231 2.332773
H 0.743738 5.175914 2.659138
H 2.490996 3.586319 1.848918
H 1.870521 1.508868 0.742107
C -3.883875 -1.247369 0.147956
C -4.654164 -0.509850 1.061805
C -5.993413 -0.814583 1.274014
C -6.595370 -1.859054 0.574575
C -5.845411 -2.593441 -0.343854
C -4.506379 -2.290172 -0.558336
H -3.953239 -2.855733 -1.301209
H -6.308509 -3.400805 -0.903222
H -7.642019 -2.096894 0.739229
H -6.570130 -0.236092 1.989928
H -4.177289 0.297960 1.605542
H -1.878475 -2.814554 -0.925962
O 1.561178 -0.681970 -3.150782
O 0.115455 -3.786036 -1.990794

1H

C -1.235982 -1.608291 -0.565826
C -2.274867 -0.825811 -0.215789

N -2.078514 0.554126 -0.094690
C -0.873891 1.163547 -0.319676
C 0.185779 0.371923 -0.658255
N 1.488561 0.687140 -0.877945
C 2.096089 -0.443196 -1.174203
C 1.195216 -1.601491 -1.142990
O 1.398894 -2.806434 -1.342219
N -0.003864 -1.020884 -0.816077
C 3.563126 -0.551564 -1.451519
C 4.344761 -0.872686 -0.189530
C 4.299008 -2.157953 0.359610
C 4.990606 -2.451069 1.531235
C 5.736991 -1.462242 2.171069
C 5.783880 -0.179026 1.632393
C 5.089383 0.113008 0.459644
H 5.128726 1.116732 0.043841
H 6.362411 0.597858 2.123875
H 6.278565 -1.691526 3.084022
H 4.947873 -3.454139 1.945951
H 3.704422 -2.921826 -0.134709
H 3.914142 0.390780 -1.881428
H 3.720581 -1.342190 -2.193379
C -0.802017 2.625695 -0.149304
C -0.033663 3.410501 -1.017100
C 0.012972 4.789550 -0.849917
C -0.705731 5.397627 0.177697
C -1.472426 4.621073 1.043314
C -1.522779 3.240928 0.881964
H -2.102306 2.639319 1.576511
H -2.027976 5.088243 1.850073
H -0.668085 6.475172 0.303555
H 0.607031 5.392107 -1.529647
H 0.518391 2.936938 -1.820571
C -3.625144 -1.340680 0.085522
C -4.383632 -0.784464 1.122693
C -5.651625 -1.280263 1.409172
C -6.173907 -2.338548 0.669485
C -5.423613 -2.897000 -0.363442
C -4.160547 -2.397027 -0.660181
H -3.592848 -2.813550 -1.486786
H -5.828268 -3.715400 -0.950612
H -7.163214 -2.724497 0.894544
H -6.227991 -0.843023 2.218416
H -3.975108 0.021946 1.725709
H -1.283976 -2.686201 -0.632411

H -2.899222 1.141918 -0.044011

3H

C -0.771889 -1.539748 -0.883525
C -1.730387 -0.807320 -0.229650
N -1.394676 0.507518 0.026432
C -0.246607 1.105779 -0.250831
C 0.753904 0.301380 -0.934292
N 1.967460 0.664958 -1.207876
C 2.596810 -0.451221 -1.860459
C 1.727303 -1.765063 -1.788091
O 1.883352 -2.860458 -1.233136
N 0.369509 -0.973091 -1.267304
C 4.087228 -0.549111 -1.586230
C 4.434791 -0.427423 -0.119832
C 3.957503 -1.360514 0.807681
C 4.270366 -1.238702 2.158371
C 5.068299 -0.185822 2.604541
C 5.548050 0.746444 1.688644
C 5.229877 0.625056 0.337068
H 5.604057 1.360106 -0.371391
H 6.169292 1.571864 2.024424
H 5.312870 -0.093009 3.658675
H 3.890036 -1.969483 2.866755
H 3.325302 -2.173938 0.459194
H 4.581947 0.245855 -2.155261
H 4.439626 -1.505084 -1.990371
C -0.037967 2.497728 0.150696
C 0.692959 3.364239 -0.674481
C 0.846626 4.695978 -0.312005
C 0.291857 5.170057 0.875188
C -0.419792 4.308336 1.707452
C -0.587098 2.977132 1.349577
H -1.107855 2.307827 2.028768
H -0.836307 4.669295 2.641757
H 0.421690 6.210069 1.156993
H 1.403197 5.365915 -0.958920
H 1.131898 2.990583 -1.591172
C -3.048386 -1.301716 0.198993
C -3.596880 -0.910228 1.426540
C -4.843025 -1.385937 1.820347
C -5.545652 -2.268642 1.003825
C -4.999507 -2.670578 -0.213411
C -3.762021 -2.184702 -0.620034
H -3.357028 -2.475274 -1.584788

H -5.544671 -3.353651 -0.856899
H -6.515575 -2.643012 1.315366
H -5.257201 -1.076404 2.774494
H -3.041720 -0.257888 2.095410
H -0.892678 -2.594266 -1.107106
O 2.330874 -0.439185 -3.289419
O 1.641045 -1.729585 -3.244741
H -2.120251 1.084269 0.447122

2'

C 1.45911 -0.74586 -0.72538
N 0.08997 -0.59510 -0.45079
C -0.87057 -1.57230 -0.44384
C -2.15690 -1.20439 -0.17602
N -2.48067 0.13109 0.01658
C -1.59414 1.09019 -0.00256
C -0.16917 0.77654 -0.22257
N 0.87094 1.53474 -0.25950
C 2.04575 0.66351 -0.50829
C -2.06544 2.47792 0.21974
C -1.26959 3.60425 -0.06195
C -1.77144 4.89123 0.13979
C -3.06483 5.07970 0.62889
C -3.86275 3.96607 0.91613
C -3.37069 2.68143 0.71241
C -3.27289 -2.17469 -0.10612
C -4.59582 -1.74127 -0.30121
C -5.65631 -2.64573 -0.25285
C -5.41869 -4.00034 -0.00649
C -4.10809 -4.44108 0.19886
C -3.04687 -3.53809 0.15575
O 1.98934 -1.79437 -1.03227
H -0.52300 -2.57342 -0.66497
H -0.26184 3.46489 -0.43120
H -1.14489 5.74948 -0.08729
H -3.44931 6.08331 0.78892
H -4.86915 4.10168 1.30268
H -3.98581 1.81677 0.93524
H -4.78131 -0.69020 -0.49402
H -6.67125 -2.29061 -0.41062
H -6.24539 -4.70415 0.03301
H -3.91126 -5.48929 0.40677
H -2.04013 -3.89580 0.35137
C 2.92061 0.69040 0.78037
H 2.28284 0.47228 1.64426

H 3.24371 1.73321 0.82974
C 4.11059 -0.23293 0.73929
C 4.19296 -1.34777 1.58336
C 5.16121 0.02006 -0.15870
C 5.30348 -2.19571 1.54293
H 3.38175 -1.55420 2.27890
C 6.26986 -0.82693 -0.19815
H 5.06175 0.89090 -0.80710

C 6.34652 -1.93746 0.65014
H 5.35279 -3.05541 2.20685
H 7.08050 -0.62319 -0.89473
H 7.21151 -2.59536 0.61541
O 2.74532 0.99739 -1.66575
O 3.49177 2.25242 -1.47303

References

- [1] Lin Jiang, Eric A Althoff, Fernando R Clemente, Lindsey Doyle, Daniela Rothlisberger, Alexandre Zanghellini, Jasmine L Gallaher, Jamie L Betker, Fujie Tanaka, Carlos F Barbas III, et al. De novo computational design of retro-aldol enzymes. *science*, 319(5868):1387–1391, 2008.
- [2] Daniela Röthlisberger, Olga Khersonsky, Andrew M Wollacott, Lin Jiang, Jason DeChancie, Jamie Betker, Jasmine L Gallaher, Eric A Althoff, Alexandre Zanghellini, Orly Dym, et al. Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192):190–195, 2008.
- [3] Hsien-Wei Yeh, Omran Karmach, Ao Ji, David Carter, Manuela M Martins-Green, and Hui-wang Ai. Red-shifted luciferase–luciferin pairs for enhanced bioluminescence imaging. *Nature methods*, 14(10):971–974, 2017.
- [4] Anna C Love and Jennifer A Prescher. Seeing (and using) the light: recent developments in bioluminescence technology. *Cell chemical biology*, 27(8):904–920, 2020.
- [5] Aisha J Syed and James C Anderson. Applications of bioluminescence in biotechnology and beyond. *Chemical Society Reviews*, 50(9):5668–5705, 2021.
- [6] Hsien-Wei Yeh and Hui-Wang Ai. Development and applications of bioluminescent and chemiluminescent reporters and biosensors. *Annual review of analytical chemistry (Palo Alto, Calif.)*, 12(1):129, 2019.
- [7] Giorgia Zambito, Chintan Chawda, and Laura Mezzanotte. Emerging tools for bioluminescence imaging. *Current Opinion in Chemical Biology*, 63:86–94, 2021.
- [8] Svetlana V Markova, Marina D Larionova, and Eugene S Vysotski. Shining light on the secreted luciferases of marine copepods: Current knowledge and applications. *Photochemistry and photobiology*, 95(3):705–721, 2019.
- [9] Tianyu Jiang, Lupei Du, and Minyong Li. Lighting up bioluminescence with coelenterazine: strategies and applications. *Photochemical & Photobiological Sciences*, 15(4):466–480, 2016.
- [10] Elisa Michelini, Luca Cevenini, Laura Mezzanotte, Danielle Ablamsky, Tara Southworth, Bruce Branchini, and Aldo Roda. Spectral-resolved gene technology for multiplexed bioluminescence and high-content screening. *Analytical chemistry*, 80(1):260–267, 2008.
- [11] Colin M Rathbun, William B Porterfield, Krysten A Jones, Marian J Sagoe, Monique R Reyes, Christine T Hua, and Jennifer A Prescher. Parallel screening for rapid identification of orthogonal bioluminescent tools. *ACS central science*, 3(12):1254–1261, 2017.
- [12] Hsien-Wei Yeh, Tianchen Wu, Minghai Chen, and Hui-wang Ai. Identification of factors complicating bioluminescence imaging. *Biochemistry*, 58(12):1689–1697, 2019.

- [13] Yichi Su, Joel R Walker, Yunhee Park, Thomas P Smith, Lan Xiang Liu, Mary P Hall, Louai Labanieh, Robin Hurst, David C Wang, Lance P Encell, et al. Novel nanoluc substrates enable bright two-population bioluminescence imaging in animals. *Nature Methods*, 17(8):852–860, 2020.
- [14] Angela Lombardi, Fabio Pirro, Ornella Maglio, Marco Chino, and William F DeGrado. De novo design of four-helix bundle metalloproteins: One scaffold, diverse reactivities. *Accounts of chemical research*, 52(5):1148–1159, 2019.
- [15] Marco Chino, Ornella Maglio, Flavia Natri, Vincenzo Pavone, William F DeGrado, and Angela Lombardi. Artificial diiron enzymes with a de novo designed four-helix bundle structure. *European Journal of Inorganic Chemistry*, 2015(21):3371–3390, 2015.
- [16] Sophie Basler, Sabine Studer, Yike Zou, Takahiro Mori, Yusuke Ota, Anna Camus, H Adrian Bunzel, Roger C Helgeson, Kendall N Houk, Gonzalo Jiménez-Osés, et al. Efficient lewis acid catalysis of an abiological reaction in a de novo protein scaffold. *Nature Chemistry*, 13(3):231–235, 2021.
- [17] Ivan Anishchenko, Samuel J Pellock, Tamuka M Chidyausiku, Theresa A Ramelot, Sergey Ovchinnikov, Jingzhou Hao, Khushboo Bafna, Christoffer Norn, Alex Kang, Asim K Bera, et al. De novo protein design by deep network hallucination. *Nature*, 600(7889):547–552, 2021.
- [18] Jue Wang, Sidney Lisanza, David Juergens, Doug Tischer, Joseph L Watson, Karla M Castro, Robert Ragotte, Amijai Saragovi, Lukas F Milles, Minkyung Baek, et al. Scaffolding protein functional sites using deep learning. *Science*, 377(6604):387–394, 2022.
- [19] Christoffer Norn, Basile IM Wicky, David Juergens, Sirui Liu, David Kim, Doug Tischer, Brian Koepnick, Ivan Anishchenko, Foldit Players, David Baker, et al. Protein sequence design by conformational landscape optimization. *Proceedings of the National Academy of Sciences*, 118(11):e2017228118, 2021.
- [20] Jianyi Yang, Ivan Anishchenko, Hahnbeom Park, Zhenling Peng, Sergey Ovchinnikov, and David Baker. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117(3):1496–1503, 2020.
- [21] Benjamin Basanta, Matthew J Bick, Asim K Bera, Christoffer Norn, Cameron M Chow, Lauren P Carter, Inna Goreshnik, Frank Dimaio, and David Baker. An enumerative algorithm for de novo design of proteins with diverse pocket structures. *Proceedings of the National Academy of Sciences*, 117(36):22135–22145, 2020.
- [22] Andreas Markus Loening, Timothy David Fenn, and Sanjiv Sam Gambhir. Crystal structures of the luciferase and green fluorescent protein from renilla reniformis. *Journal of molecular biology*, 374(4):1017–1028, 2007.
- [23] Yuri Tomabechi, Takamitsu Hosoya, Haruhiko Ehara, Shun-ichi Sekine, Mikako Shirouzu, and Satoshi Inouye. Crystal structure of nanokaz: The mutated 19 kda component of oplophorus luciferase catalyzing the bioluminescent reaction with coelenterazine. *Biochemical and biophysical research communications*, 470(1):88–93, 2016.

- [24] Nan Wu, Naohiro Kobayashi, Kengo Tsuda, Satoru Unzai, Tomonori Saotome, Yutaka Kuroda, and Toshio Yamazaki. Solution structure of gaussian luciferase with five disulfide bonds and identification of a putative coelenterazine binding cavity by heteronuclear nmr. *Scientific reports*, 10(1):1–11, 2020.
- [25] Bo-Wen Ding and Ya-Jun Liu. Bioluminescence of firefly squid via mechanism of single electron-transfer oxygenation and charge-transfer-induced luminescence. *Journal of the American Chemical Society*, 139(3):1106–1119, 2017.
- [26] Hiroshi Isobe, Syusuke Yamanaka, Seiki Kuramitsu, and Kizashi Yamaguchi. Regulation mechanism of spin-orbit coupling in charge-transfer-induced luminescence of imidazopyrazinone derivatives. *Journal of the American Chemical Society*, 130(1):132–149, 2008.
- [27] Hiroyuki Kondo, Takayuki Igarashi, Shojiro Maki, Haruki Niwa, Hiroshi Ikeda, and Takashi Hirano. Substituent effects on the kinetics for the chemiluminescence reaction of 6-arylimidazo [1, 2-a] pyrazin-3 (7h)-ones (cypridina luciferin analogues): support for the single electron transfer (set)-oxygenation mechanism with triplet molecular oxygen. *Tetrahedron letters*, 46(45):7701–7704, 2005.
- [28] Bruce R Branchini, Curran E Behney, Tara L Southworth, Danielle M Fontaine, Andrew M Gulick, David J Vinyard, and Gary W Brudvig. Experimental support for a single electron-transfer oxidation mechanism in firefly bioluminescence. *Journal of the American Chemical Society*, 137(24):7592–7595, 2015.
- [29] Jiayi Dou, Anastassia A Vorobieva, William Sheffler, Lindsey A Doyle, Hahnbeom Park, Matthew J Bick, Binchen Mao, Glenna W Foight, Min Yen Lee, Lauren A Gagnon, et al. De novo design of a fluorescence-activating β -barrel. *Nature*, 561(7724):485–491, 2018.
- [30] Longxing Cao, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Joon Sung Park, Kevin M Jude, Iva Marković, Rameshwar U Kadam, Koen HG Verschueren, et al. Design of protein-binding proteins from the target structure alone. *Nature*, 605(7910):551–560, 2022.
- [31] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.
- [32] Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning-based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.
- [33] Hsien-Wei Yeh, Ying Xiong, Tianchen Wu, Minghai Chen, Ao Ji, Xinyu Li, and Hui-wang Ai. Atp-independent bioluminescent reporter variants to improve in vivo imaging. *ACS chemical biology*, 14(5):959–965, 2019.

- [34] Ying Xiong, Yiyu Zhang, Zefan Li, Md Shamim Reza, Xinyu Li, Xiaodong Tian, and Huiwang Ai. An engineered amber-emitting nano luciferase and its use for immunobioluminescence imaging in vivo. *bioRxiv*, 2022.
- [35] S Bhaumik and SS Gambhir. Optical imaging of renilla luciferase reporter gene expression in living mice. *Proceedings of the National Academy of Sciences*, 99(1):377–382, 2002.
- [36] Christopher Szent-Gyorgyi, Byron T Ballou, Erich Dagnal, and Bruce Bryan. Cloning and characterization of new bioluminescent proteins. In *Biomedical Imaging: Reporters, Dyes, and Instrumentation*, volume 3600, page 4–11. SPIE, 1999.
- [37] Mary P Hall, James Unch, Brock F Binkowski, Michael P Valley, Braeden L Butler, Monika G Wood, Paul Otto, Kristopher Zimmerman, Gediminas Vidugiris, Thomas Machleidt, et al. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS chemical biology*, 7(11):1848–1857, 2012.
- [38] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R Dustin Schaeffer, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.
- [39] BIM Wicky, LF Milles, A Courbet, RJ Ragotte, J Dauparas, E Kinfu, S Tipps, RD Kibler, M Baek, F DiMaio, et al. Hallucinating symmetric protein assemblies. *Science*, 378(6615):56–61, 2022.
- [40] Lars Giger, Sami Caner, Richard Obexer, Peter Kast, David Baker, Nenad Ban, and Donald Hilvert. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nature chemical biology*, 9(8):494–498, 2013.
- [41] Caroline K Brennan, Zi Yao, Lorenzo Scipioni, Hongtao Chen, Kevin Ng, Giulia Tedeschi, Michelle A Digman, and Jennifer A Prescher. Multiplexed bioluminescence microscopy via phasor analysis. *Biophysical Journal*, 121(3):142a, 2022.
- [42] Yoriko Ando, Kazuki Niwa, Nobuyuki Yamada, Tsutomu Irie, Toshiteru Enomoto, Hidehiro Kubota, Yoshihiro Ohmiya, and Hidefumi Akiyama. Development of a quantitative bio/chemiluminescence spectrometer determining quantum yields: Re-examination of the aqueous luminol chemiluminescence standard. *Photochemistry and photobiology*, 83(5):1205–1210, 2007.
- [43] Jianyi Yang, Ivan Anishchenko, Hahnbeom Park, Zhenling Peng, Sergey Ovchinnikov, and David Baker. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117(3):1496–1503, 2020.
- [44] Weizhong Li and Adam Godzik. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659, 2006.
- [45] Daniel P Farrell, Ivan Anishchenko, Shabih Shakeel, Anna Lauko, Lori A Passmore, David Baker, and Frank DiMaio. Deep learning enables the atomic structure determination of the fanconi anemia core complex from cryoem. *IUCrJ*, 7(5):881–892, 2020.

- [46] Jason C Klein, Marc J Lajoie, Jerrod J Schwartz, Eva-Maria Strauch, Jorgen Nelson, David Baker, and Jay Shendure. Multiplex pairwise assembly of array-derived dna oligonucleotides. *Nucleic acids research*, 44(5):e43–e43, 2016.
- [47] Yang Zhang and Jeffrey Skolnick. Tm-align: a protein structure alignment algorithm based on the tm-score. *Nucleic acids research*, 33(7):2302–2309, 2005.
- [48] Saul B Needleman and Christian D Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3):443–453, 1970.
- [49] Toshiyuki Oda, Kyungtaek Lim, and Kentaro Tomii. Simple adjustment of the sequence weight algorithm remarkably enhances psi-blast performance. *BMC bioinformatics*, 18(1):1–8, 2017.
- [50] Stephen F Altschul, Thomas L Madden, Alejandro A Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J Lipman. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic acids research*, 25(17):3389–3402, 1997.
- [51] Brian Coventry and David Baker. Protein sequence optimization with a pairwise decomposable penalty for buried unsatisfied hydrogen bonds. *PLoS computational biology*, 17(3):e1008061, 2021.
- [52] Adam JT Smith, Roger Muller, Miguel D Toscano, Peter Kast, Homme W Hellinga, Donald Hilvert, and KN Houk. Structural reorganization and preorganization in enzyme active sites: comparisons of experimental and theoretically ideal active site geometries in the multistep serine esterase reaction cycle. *Journal of the American Chemical Society*, 130(46):15361–15373, 2008.
- [53] Elizabeth H Kellogg, Andrew Leaver-Fay, and David Baker. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins: Structure, Function, and Bioinformatics*, 79(3):830–838, 2011.
- [54] Hahnbeom Park, Philip Bradley, Per Greisen Jr, Yuan Liu, Vikram Khipple Mulligan, David E Kim, David Baker, and Frank DiMaio. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *Journal of chemical theory and computation*, 12(12):6201–6212, 2016.
- [55] Jiashu Liang, Xintian Feng, Diptarka Hait, and Martin Head-Gordon. Revisiting the performance of time-dependent density functional theory for electronic excitations: Assessment of 43 popular and recently developed functionals from rungs one to four. *Journal of Chemical Theory and Computation*, 2022.
- [56] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics*, 10(44):6615–6620, 2008.

- [57] RHWJ Ditchfield, W J_ Hehre, and John A Pople. Self-consistent molecular-orbital methods. ix. an extended gaussian-type basis for molecular-orbital studies of organic molecules. *The Journal of Chemical Physics*, 54(2):724–728, 1971.
- [58] Stefan Grimme. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *Journal of chemical theory and computation*, 15(5):2847–2862, 2019.
- [59] Philipp Pracht, Fabian Bohle, and Stefan Grimme. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical Chemistry Chemical Physics*, 22(14):7169–7192, 2020.
- [60] Guilian Luchini, J Alegre-Requena, Ignacio Funes-Ardoiz, and Robert S Paton. Good-vibes: automated thermochemistry for heterogeneous computational chemistry data. *F1000Research*, 9, 2020.
- [61] Yi-Pei Li, Joseph Gomes, Shaama Mallikarjun Sharada, Alexis T Bell, and Martin Head-Gordon. Improved force-field parameters for qm/mm simulations of the energies of adsorption for molecules in zeolites and a free rotor correction to the rigid rotor harmonic oscillator model for adsorption enthalpies. *The Journal of Physical Chemistry C*, 119(4):1840–1850, 2015.
- [62] Romelia Salomon-Ferrer, Andreas W Gotz, Duncan Poole, Scott Le Grand, and Ross C Walker. Routine microsecond molecular dynamics simulations with amber on gpus. 2. explicit solvent particle mesh ewald. *Journal of chemical theory and computation*, 9(9):3878–3888, 2013.
- [63] AD Becke. Density-functional thermochemistry. iii the role of exact exchange j chem phys 98: 5648–5652, 1993.
- [64] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of chemical physics*, 132(15):154104, 2010.
- [65] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. Effect of the damping function in dispersion corrected density functional theory. *Journal of computational chemistry*, 32(7):1456–1465, 2011.
- [66] Jens Meiler and David Baker. Rosettaligand: Protein–small molecule docking with full side-chain flexibility. *Proteins: Structure, Function, and Bioinformatics*, 65(3):538–548, 2006.
- [67] Ian W Davis and David Baker. Rosettaligand docking with full ligand and receptor flexibility. *Journal of molecular biology*, 385(2):381–392, 2009.
- [68] Ian W Davis, Kaushik Raha, Martha S Head, and David Baker. Blind docking of pharmaceutically relevant compounds using rosettaligand. *Protein science*, 18(9):1998–2002, 2009.

- [69] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.
- [70] Christopher I Bayly, Piotr Cieplak, Wendy Cornell, and Peter A Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993.
- [71] Brent H Besler, Kenneth M Merz Jr, and Peter A Kollman. Atomic charges derived from semiempirical methods. *Journal of computational chemistry*, 11(4):431–439, 1990.
- [72] U Chandra Singh and Peter A Kollman. An approach to computing electrostatic charges for molecules. *Journal of computational chemistry*, 5(2):129–145, 1984.
- [73] William L Jorgensen, Jayaraman Chandrasekhar, Jeffrey D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.
- [74] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.
- [75] Darrin M York, Alexander Wlodawer, Lee G Pedersen, and Tom A Darden. Atomic-level accuracy in simulations of large protein crystals. *Proceedings of the National Academy of Sciences*, 91(18):8715–8718, 1994.
- [76] Daniel R Roe and Thomas E Cheatham III. Ptraj and cpptraj: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation*, 9(7):3084–3095, 2013.

Conclusion

Here, we presented several examples of engineered catalysts along with computational physical models that enhance our mechanistic understanding of how these catalysts work. In Chapter 2, we discussed a dirhodium catalyst developed for an enantioselective Si-H insertion and showed that the chiral environment of the catalyst limited conformations of the transition state, only allowing the observed product. In Chapter 3, we investigated a family of laboratory evolved enzymes and showed that a catalytic lysine acts as a general acid to promote chlorination via HOCl. Mutations in the active site altered substrate binding and ultimately led to the differences in site selectivity. Finally, in Chapter 4, we investigated a set of computationally designed enzymes and showed that the most active mutant was best at stabilizing the reaction intermediates. The mechanistic insight gained from these is now being used to engineer a new generation of catalysts with improved activity.

Dirhodium catalysts are among the most powerful and widely used in chemistry. They are used extensively in C-H insertion [1], Si-H insertion [2] and cyclopropanation [3, 4]. Recent work from Lewis and coworkers has also led to artificial metalloenzyme that contain a dirhodium cofactor capable of catalyzing enantioselective cyclopropanation [5]. In Chapter 2, we saw that the power in these catalysts arises from their ability to stabilize the highly reactive carbene. The rest of the catalyst acted as a steric pocket to control the selectivity of the reaction. DFT calculations using the tetraformate catalyst, which lacked any steric environment, showed that both enantiomeric transition states were equal in energy. Electronic differences in the aryl rings of the substrate limited the conformation of these transition states to favor the electron-rich aryl conjugated to

the vacant *p*-orbital of the carbene. In the environment of (*S*)-C-1, these electronic and steric constraints allowed only the transition state that led to the observed product.

Despite the power of these catalysts, they are difficult to scale for a several reasons; chemical synthesis of the equatorial ligands can be challenging or expensive, and rhodium is a rare and extremely expensive metal. Increasing efforts have been made to address these issues by use of biocatalysis. The impressive heme-containing enzymes evolved by Arnold and coworkers [6] demonstrate that the same reactivity is possible using only iron and the 20 canonical amino acids. In these enzymes, the iron heme acts as the dirhodium core to stabilize the carbene or nitrene, and the rest of the enzyme imparts selectivity onto the reaction. No doubt, the field of chemical catalysis will be dominated by enzymes in the near future.

Directed evolution is currently the state of the art in enzyme engineering and has resulted in the most effective engineered enzymes to date. The relatively simple idea to imitate the natural process of evolution does not rely on any structural or mechanistic understanding of the enzyme, bypassing an enormous bottleneck of rational design. However, additional insight from experiments [7] or machine learning [8] have been able to improve this process, indicating that mechanistic insight can still be beneficial. In Chapter 3, we investigated a family of mutants evolved to have orthogonal selectivity for the chlorination of tryptamine. Dramatic differences in selectivity arising from only a handful of mutations made this an ideal system to understand the role of individual mutations. Ultimately, it was clear that differences in selectivity were due to differences in substrate binding. DFT calculations showed the intrinsic differences in substrate reactivity, and MD simulations showed how each enzyme was able to influence the reaction through substrate binding.

The physical model we developed in Chapter 3 was used to predict selectivity on nonnative substrates with high accuracy. DFT-computed transition state and Wheland intermediate energies were the best predictor of site selectivity, indicating that the enzymes have little catalyst control over the substrates they were not evolved for. In these cases where substrate binding is low, selectivity is dominated by substrate control. On this principle, new enzymes can be engineered that optimize substrate binding in a desired orientation. New protein design methods that predict

mutations to stabilize a desired protein fold [9, 10] will likely be increasingly used to this end. As the success rate of these computational designs increases, rational design will become increasingly popular.

Computational protein design has benefited greatly from recent advancements in machine learning. Rapid, accurate predictions of protein structure enable high throughput computational design methods. Some of these networks have also been reversed to design novel stable protein backbones [11, 12, 13]. In Chapter 4, we investigate a set of proteins that were generated using the family-wide hallucination approach. This method uses the trRosetta network [14] to generate protein sequences that contain a given fold topology. Theozymes that bind to and stabilize the forming anion were then docked into each of these scaffolds using the RIFdock [15] algorithm before a final sequence optimization.

Following an initial round of computational designs, the most active enzymes were generated using site-saturation mutagenesis. The most active enzyme, LuxSit-i, contained only 3 mutations relative to the initial design and resulted in over 100-fold improvement. Again, we benefited from having a set of proteins that were all structurally similar, yet with a small number of mutations displayed a wide range of activity. DFT calculations confirmed that stabilization of the forming anion was necessary for the reaction, and MD simulations showed that the most active enzyme was the best at facilitating this stabilization. Additionally, we showed that the binding pocket of the most active enzyme was unable to accept substrates with benzylic substituents at R⁸, explaining the selectivity differences of this enzyme.

Importantly, we also showed that the most active enzyme was the best at stabilizing the *transition state* rather than the ground state. Changing the design target to bind **2'** rather than **1** will likely lead to higher success rates in the future. Incorporation of transition state models into enzyme design will likely have an enormous impact on the success rates in general. This, in addition to improved structure prediction networks and networks that accommodate protein-ligand interactions, should make enzyme design somewhat commonplace in the coming years.

It seems inevitable at this point that enzyme engineering, particularly enzyme design, will

dominate chemical catalysis in the foreseeable future. Statistical models in the form of machine learning are the driving force behind this revolution, but physical models will still be critical for successful designs. For example, protein inpainting [16], which uses a structure prediction neural network to design proteins with a functional site still requires an accurate physical model of the target as an input. DFT calculations are still needed to determine the ideal arrangements of active sites. Molecular mechanics will also be needed in places where DFT is impractical, however the role of these calculations will diminish as computing power improves. Nevertheless, MD simulations are still critical for understanding the subtle differences between similar enzymes, as we noted here. In total, no one method can be used to reliably engineer a functional catalyst and it requires extensive knowledge of each process to arrive at the desired product. Catalyst engineering has far to go before it is routinely adopted in labs across the world, however the potential impact of this field is enormous and could change the way chemistry and biology are conducted.

References

- [1] Huw ML Davies and Kuangbiao Liao. Dirhodium tetracarboxylates as catalysts for selective intermolecular c–h functionalization. *Nature Reviews Chemistry*, 3(6):347–360, 2019.
- [2] Hoda Keipour, Virginie Carreras, and Thierry Ollevier. Recent progress in the catalytic carbene insertion reactions into the silicon–hydrogen bond. *Organic & Biomolecular Chemistry*, 15(26):5441–5456, 2017.
- [3] Frady G Adly, Michael G Gardiner, and Ashraf Ghanem. Design and synthesis of novel chiral dirhodium (ii) carboxylate complexes for asymmetric cyclopropanation reactions. *Chemistry–A European Journal*, 22(10):3447–3461, 2016.
- [4] Frady G Adly, Michael G Gardiner, and Ashraf Ghanem. Design and synthesis of novel chiral dirhodium (ii) carboxylate complexes for asymmetric cyclopropanation reactions. *Chemistry–A European Journal*, 22(10):3447–3461, 2016.
- [5] Poonam Srivastava, Hao Yang, Ken Ellis-Guardiola, and Jared C Lewis. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nature communications*, 6(1):1–8, 2015.
- [6] Yang Yang and Frances H. Arnold. Navigating the unnatural reaction space: Directed evolution of heme proteins for selective carbene and nitrene transfer. *Accounts of Chemical Research*, 54(5):1209–1225, 2021. PMID: 33491448.
- [7] Sagar Bhattacharya, Eleonora G Margheritis, Katsuya Takahashi, Alona Kulesha, Areetha D’Souza, Inhye Kim, Jennifer H Yoon, Jeremy RH Tame, Alexander N Volkov, Olga V Makhlynets, et al. Nmr-guided directed evolution. *Nature*, 610(7931):389–393, 2022.
- [8] Kevin K Yang, Zachary Wu, and Frances H Arnold. Machine-learning-guided directed evolution for protein engineering. *Nature methods*, 16(8):687–694, 2019.
- [9] Namrata Anand, Raphael Eguchi, Irimpan I. Mathews, Carla P. Perez, Alexander Derry, Russ B. Altman, and Po-Ssu Huang. Protein sequence design with a learned potential. *Nature Communications*, 13(1):746, Feb 2022.
- [10] J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker. Robust deep learning based protein sequence design using proteinmpnn. *bioRxiv*, 2022.
- [11] Ivan Anishchenko, Samuel J. Pellock, Tamuka M. Chidyausiku, Theresa A. Ramelot, Sergey Ovchinnikov, Jingzhou Hao, Khushboo Bafna, Christoffer Norm, Alex Kang, Asim K. Bera, Frank DiMaio, Lauren Carter, Cameron M. Chow, Gaetano T. Montelione, and David Baker. De novo protein design by deep network hallucination. *Nature*, 600(7889):547–552, Dec 2021.

- [12] Jue Wang, Sidney Lisanza, David Juergens, Doug Tischer, Joseph L. Watson, Karla M. Castro, Robert Ragotte, Amijai Saragovi, Lukas F. Milles, Minkyung Baek, Ivan Anishchenko, Wei Yang, Derrick R. Hicks, Marc Expòsit, Thomas Schlichthaerle, Jung-Ho Chun, Justas Dauparas, Nathaniel Bennett, Basile I. M. Wicky, Andrew Muenks, Frank DiMaio, Bruno Correia, Sergey Ovchinnikov, and David Baker. Scaffolding protein functional sites using deep learning. *Science*, 377(6604):387–394, 2022.
- [13] BIM Wicky, LF Milles, A Courbet, RJ Ragotte, J Dauparas, E Kinfu, S Tipps, RD Kibler, M Baek, F DiMaio, et al. Hallucinating symmetric protein assemblies. *Science*, 378(6615):56–61, 2022.
- [14] Jianyi Yang, Ivan Anishchenko, Hahnbeom Park, Zhenling Peng, Sergey Ovchinnikov, and David Baker. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117(3):1496–1503, 2020.
- [15] Jiayi Dou, Anastassia A. Vorobieva, William Sheffler, Lindsey A. Doyle, Hahnbeom Park, Matthew J. Bick, Binchen Mao, Glenna W. Foight, Min Yen Lee, Lauren A. Gagnon, Lauren Carter, Banumathi Sankaran, Sergey Ovchinnikov, Enrique Marcos, Po-Ssu Huang, Joshua C. Vaughan, Barry L. Stoddard, and David Baker. De novo design of a fluorescence-activating β -barrel. *Nature*, 561(7724):485–491, Sep 2018.
- [16] Jue Wang, Sidney Lisanza, David Juergens, Doug Tischer, Joseph L. Watson, Karla M. Castro, Robert Ragotte, Amijai Saragovi, Lukas F. Milles, Minkyung Baek, Ivan Anishchenko, Wei Yang, Derrick R. Hicks, Marc Expòsit, Thomas Schlichthaerle, Jung-Ho Chun, Justas Dauparas, Nathaniel Bennett, Basile I. M. Wicky, Andrew Muenks, Frank DiMaio, Bruno Correia, Sergey Ovchinnikov, and David Baker. Scaffolding protein functional sites using deep learning. *Science*, 377(6604):387–394, 2022.