Computational Methods for Modifying Enzyme Specificity:

from Molding the Active Site to Allosteric Considerations

by

Bethany Kolbaba Kartchner

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2023 by the
Graduate Supervisory Committee:

Jeremy H. Mills, Chair
Giovanna Ghirlanda
Wade D. Van Horn

ARIZONA STATE UNIVERSITY

December 2023

ABSTRACT

Enzymes keep life nicely humming along by catalyzing important reactions at relevant timescales. Despite their immediate importance, how enzymes recognize and bind their substrate in a sea of cytosolic small molecules, carry out the reaction, and release their product in microseconds is still relatively opaque. Methods to elucidate enzyme substrate specificity indicate that the shape of the active site and the amino acid residues therein play a major role. However, lessons from Directed Evolution experiments reveal the importance of residues far from the active site in modulating substrate specificity. Enzymes are dynamic macromolecules composed of networks of interactions integrating the active site, where the chemistry occurs, to the rest of the protein. The objective of this work is to develop computational methods to modify enzyme ligand specificity, either through molding the active site to accommodate a novel ligand, or by identifying distal mutations that can allosterically alter specificity.

To this end, two homologues in the β-lactamase family of enzymes, TEM-1, and an ancestrally reconstructed variant, GNCA, were studied to identify whether the modulation of position-specific distal-residue flexibility could modify ligand specificity. RosettaDesign was used to create TEM-1 variants with altered dynamic patterns. Experimental characterization of ten designed proteins indicated that mutations to residues surrounding rigid, highly coupled residues substantially affected both enzymatic activity and stability. In contrast, native-like activities and stabilities were maintained when flexible, uncoupled residues, were targeted. Five of the TEM-1 variants were crystallized to see if the changes in function observed were due to architectural changes to the active site. In a second project, a computational platform using RosettaDesign was

developed to remodel the firefly luciferase active site to accommodate novel luciferins. This platform resulted in the development of five luciferin-luciferase pairs with red-shifted emission maxima, ready for multicomponent bioluminescent imaging applications in tissues. Although the projects from this work focus on two classes of proteins, they provide insight into the structure-function relationship of ligand specificity in enzymes and are broadly applicable to other systems.

ACKNOWLEDGMENTS

It not only takes a village to raise a child, but it takes a village to develop a scientist. This work may have been done by my hands, heart, and head, but it would never have come to fruition without the multitude of visible and invisible hands figuratively holding me up, cheering me on, and lending an ear.

First, I acknowledge the many scientific mentors who have guided me on my path. Dr. John Zikopoulos taught me Organic Chemistry, introduced me to the joys of research, and has been a great support for many years. Dr. Marcia Levitus, my first mentor at Arizona State University, who, recognizing something in me, paved the way for me to find a position as an undergraduate in the Mills lab, provided ongoing advice, and over the years has become a dear friend. Dr. J. Nathan Henderson taught me everything I know about crystallography and worked endlessly with me as I solved crystal structures. Dr. Patrick R. Gleason taught me the ins-and-outs of molecular cloning and protein purification. My committee members, Dr. Wade Van Horn and Dr. Giovanna Ghirlanda, have been instrumental in asking the right probing questions and guiding my projects.

Second, I acknowledge the many collaborator-mentors that I've had. All my projects have been collaborations with labs both at ASU and outside ASU. Dr. S. Banu Ozkan, Dr. Ismail Can Kazan, Dr. Jennifer Prescher, Dr. Anna Love, and Dr. Martin Schnermann, have imparted their own unique ways of doing science and of approaching challenges and they have been a great pleasure to work with.

Third, I acknowledge the mentors that I encountered during the journey. Dr. Elisabeth Narayanan, my manager at Moderna, gave me the space to learn coding and (big) data analysis while also building my confidence as a computationalist. Dr. Jane Clarke, whom I met at Summer RosettaCON 2019, changed my life with her advice on effectively navigating the parenting-science time balance and has been a great source of practical wisdom over the years. I am also grateful for the friends that I made working in the Mills lab, while serving on the Rosetta Commons JEDI committee, the SMS Graduate Student Council, and in the SMS program itself. They have made my scientific life sweeter.

I couldn't have asked for a better advisor than Dr. Jeremy Mills. He has not only supported me personally as I navigated the challenges of raising children while completing my doctorate, but he's molded me into the scientist I am today. His insistence on creating a good figure for a presentation, to his precise wordsmithing of my writing, to his analysis while looking at Rosetta designs have taught me how to think scientifically. Additionally, he provided me with projects to broaden my skillset which has been invaluable to me.

As I've been mentored, I've had the opportunity to mentor gifted graduate, undergraduate, and high school students in the lab: Ngoc Huynh, Jessica Williams, Ian Oross, Stephanie Donovan, Brooke Lovell, Maxwell Igwe, Michael Lau, Sneha Lakamsani, Elyssa Kartchner, and Cameron Kartchner have enriched my life as I learned as much, if not more, than I taught them. Without my graduate and undergraduate mentees, I never would have been as productive as they carried on my research as I completed my internship and traveled to conferences.

I would be remiss if I didn't mention the many people outside the lab who have a significant impact on the quality of my life. First, my husband, Todd, kept me going when I was especially discouraged or overwhelmed. I know that I wouldn't have gotten to this point without him. He supported me through the late nights, early mornings, and everything in between. He has been the perfect partner during this long road, and I adore him. Second, my children, Danika, Zoey, Cameron, Elyssa, Quinn, and Wyatt, whom I think are simply amazing, have been with me every step of the way. Every one of them has waited patiently in my office for me to finish an experiment and every one of them has done everything they could to support me. I was lucky to work with three of them in the lab itself and I hope that they treasure those memories as much as I do. All six children are growing up to be intelligent, kind, thoughtful people. I am incredibly proud of them and of all they've already accomplished in life. Third, I'm also grateful for my brothers and sisters (and their spouses who are also my brothers and sisters). They have lightened my life with fun and companionship. Finally, my parents, Dr. Scott and Joan Kolbaba, who taught me that with work I could accomplish anything I wanted to. I think that this degree is evidence that they were right. For that I'm so grateful.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 A Brief History of Enzymes

The word *enzyme* was first used in a biochemical context by Wilhelm Kühne in 1878 to describe the catalytic activity present in yeast extracts (Sumner & Somers, 2014). Prior to this nomenclature assignment, microorganisms had been used by humans for thousands of years for fermentation and cheese-making, but the nature of the molecule performing the chemical reaction was unknown and would remain unknown until 1946. Nobel Prize winning work by James B. Sumner (prize share ½), Wendell M. Stanley (prize share ¼), and John B. Northrop (prize share ¼) on the crystallization of urease, pepsin, trypsin and chymotrypsin, revealed that enzymes were made of amino acids (Nobel Foundation, 1964). We now know that enzymes are protein catalysts that play an integral role in metabolism, cell signaling, immunity, and gene expression, to name a few. Without enzymes, reactions that keep life nicely humming in a cell would take years or potentially even longer to occur and life on the timescale that we are used to would cease to exist.

Nature has naturally evolved enzymes to perform chemistries important to life in the organism housing them. These enzymes vary in catalytic ability from one of the slowest, but most abundant enzymes, rubisco that turns over 3 molecules of carbon dioxide every second (Bar-On & Milo, 2019); to one of the fastest enzymes, carbonic anhydrase that can turn over 10^6 carbon dioxide molecules a second (Lindskog & Colemant, 1973). Most enzyme efficiencies fall in between these two extremes. Since nature only optimizes enzymes to be as stable and catalytically efficient as natural

selection dictates, most native proteins do not naturally possess functions important to human applications. Thus, the field of protein engineering is built around the desire to modify or create proteins with functions important to human lives.

1.2 Methods in Protein Engineering

Protein engineering, the process of altering protein sequences to provide variants with new or enhanced properties, was born in the 1960s when Christian Anfinsen (Nobel Prize in Chemistry in 1972) demonstrated that a protein's amino acid sequence alone determines its structure which in turn determines its function (Forsen, 1993). Initially, changes to protein sequence were performed on an organismal scale through the process of stain optimization (Heckmann & Paradisi, 2020). In this technique, entire organisms were placed in mutation-inducing conditions like radiation or chemical agents. The resulting strains carrying random mutations were then screened for beneficial phenotypes. Although effective in specific cases like in the development of penicillin acylase by Bayer, (Buchholz, 2016) strain optimization was time-intensive and only applicable to organisms with short replication cycles. In the 1970-1980s improvements in DNA cloning technologies, the judicial use of restriction enzymes, and an increased understanding of protein function reduced the scale of protein engineering from the cellular to the macromolecular level (Mullis et al., 1994). Researchers were able to directly and specifically target genes encoding proteins leading to changes in function (Hughes, 2011). In 1993, Frances Arnold employed these advances to successfully use error-prone PCR to create large libraries of variants of subtilisin E with the goal of identifying mutants that could tolerate high concentrations of dimethylformamide (K. Chen et al., 1991; K. Chen & Arnold, 1993). From this point forward, the field of protein

engineering has experienced explosive growth. Its importance in medicine, agriculture, and industry was publicly acknowledged when Francis Arnold (prize share 1/3), was awarded the 2018 Nobel Prize in Chemistry for her work on engineering enzymes with novel functions (Garcia, 2018).

In the field of protein engineering, there are two principal approaches: perturbation-based strategies known as "directed evolution," and design-centric strategies based on rational, knowledge-based modifications of protein structure (Clark & Pazdernik, 2008; Tiwari et al., 2012). Directed evolution, the experimental method Frances Arnold perfected, is an approach in which a target gene is randomly mutated and partnered with a screening or selection protocol that identifies variants possessing the desired new function (F. H. Arnold et al., 2001). Rational design, the second main approach, is a site-specific, structure-based method in which portions of the protein are targeted for mutagenesis (Song et al., 2023). Novel functionality is identified using screening or genetic selections. As these approaches are not mutually exclusive, many researchers rely on an iterative combination of the two to optimize the final product (Chica et al., 2005).

1.3 Computational Design of Proteins using the Rosetta Software Suite

Despite the successes achieved by incorporating directed evolution approaches, large selections and screening procedures are cumbersome and expensive. To circumvent the pitfalls of this methodology, rational designers incorporate computational methods into the design process including the groundbreaking Rosetta Software suite for macromolecular modeling developed in the David Baker lab at the University of Washington (Bender et al., 2016; Richter et al., 2011). Rosetta is used to quickly test

hypotheses based on initial structural models that are labor intensive to perform experimentally. Algorithms for a variety of protein modeling design tasks have been developed within the Rosetta software, and the majority rely on a Monte Carlo simulated annealing search which is an iterative sampling of the sequence and/or conformational space of a protein (Rohl et al., 2004). The total energy of the modified protein is assessed using a physics and knowledge-based score function composed of 19 different weighted terms derived largely from information gleaned from the more than 200,000 crystal structures in the Protein Data Bank (PDB) (Alford et al., 2017a). Many of the score terms were developed using statistical occurrences of pair-wise interactions from crystal structures including electrostatics, the 6-12 Lennard-Jones potential for van der Waals forces, a geometric-centric hydrogen bonding potential, solvation effects, and amino acid side-chain conformation energetics (Dunbrack, Roland L., Jr. and Karplus, 1993; Kortemme et al., 2003; Lazaridis, 2003). To avoid a progress-halting descent into a local minimum on the conformational landscape, a Metropolis criterion is also applied that randomly accepts certain mutations that increase the energy (Kaufmann et al., 2010). The solutions proposed by Rosetta are predicted to be designs that will fold properly into functional, soluble proteins. However, experimental characterization of designed proteins is an integral component to a comprehensive research strategy.

Significant advances in protein engineering made possible by Rosetta include the design of proteins with the ability to bind metals and small molecules (Mills et al., 2013, 2016; Tinberg & Khare., 2017) and *de novo* design of proteins with folds that do not exist in nature (Huang et al., 2016) to name a few. Another important application for which Rosetta has been used is enzyme design (Richter et al., 2011). Novel enzymes with their

unparalleled ability to catalyze reactions under mild conditions with stereo-selectivity make them profoundly important to sustainable manufacturing processes, the creation of new small molecule therapeutics, and even toxic waste clean-up (Singh et al., 2016).

1.4 Computational Design of Enzymes: Successes and Current Challenges

Traditionally, computational design of enzymes follows the Linus Pauling view of enzymatic catalysis which is that the enzyme binds with higher affinity to the transition state (Pauling, 1946). Therefore, the design methodology is heavily weighted on the principle of transition state stabilization (Bolon & Mayo, 2001; Privett et al., 2012; Tantillo et al., 1998; Zhu & Lai, 2009). For modeling, a theoretical transition state structure, a "theozyme," consisting of the catalytic residues and a bound substrate molecule, is generated *in silico*. Ideally, the theozyme includes activation-barrier-reducing interactions between the catalytic residues and the substrate. The next step is to search a library of structures to identify a protein backbone that can accommodate the theozyme. This scaffold set can consist of crystal structure models of thermostable proteins in the PDB or, more recently, can be a *de novo* protein backbone generated by machine learning techniques like RFDiffusion (Watson et al., 2023; Yeh et al., 2023). When the theozyme is modeled into the chosen scaffold, further structural modifications to stabilize the active site by buttressing the catalytic residues are undertaken. The goal of the final computational optimizations is to design a well-packed, well-organized catalytic pocket (Smith et al., 2008). The goal of this three-step procedure is a protein that will fold into a soluble enzyme with the ability to carry out the desired reaction. This protocol has been implemented to successfully design a Kemp eliminase, a retroaldolase, ester

hydrolase, a Diels-Alderase, and more recently a luciferase (Jiang et al., 2008; Richter et al., 2012; Röthlisberger et al., 2008; Siegel et al., 2010; Yeh et al., 2023).

Despite significant advances in successfully engineering proteins with novel function, computational design methods are plagued by enzymes with very low efficiencies when compared to naturally evolved enzymes (Mak & Siegel, 2014). Hypotheses regarding the observed discrepancy in catalytic efficiencies between naturally occurring and designed enzymes include suboptimal transition state modeling, suboptimal active site access to solvent, and the inability to model productive protein dynamics into the scaffold (Chovancova et al., 2012; Kiss et al., 2010; Leaver-Fay, Jacak, et al., 2011; Mak & Siegel, 2014; Radzika & Wolfenden, 1995). Work is ongoing to tease out what exactly current enzyme design methods are missing. However, the hypothesis of greatest interest to this work is that enzyme design is not currently able to recapitulate integral dynamic movements that are key to productive enzymatic function.

The relationship between enzyme catalysis and dynamics has been a fiercely debated topic for many years (Kamerlin & Warshel, 2010; Schneider et al., 2021; Schnell et al., 2004; Warshel et al., 2006; Warshel & Bora, 2016). Understanding the relationship between structure and dynamics in enzymes is fraught with challenges because even small changes in enzyme structure can lead to changes in dynamics (Schnell et al., 2004). However, new insights into enzyme function have given weight to the importance of dynamics in enzyme function (Baker, 2010; K. Henzler-Wildman & Kern, 2007; Korendovych, Ivan V., DeGrado, 2014; Markin et al., 2021; Vaissier Welborn & Head-Gordon, 2019; Wolf-Watz et al., 2004a).

1.5 Protein Dynamics and Their Role in Enzyme Function

Early on, proteins were perceived as static entities with fixed structures. Initial models of enzymes followed in this tradition with work in 1894 by Emil Fischer on the specificity of enzymes for their substrates leading to the “lock and key" hypothesis. This model provided the first mental image of molecular recognition as a stiff interaction between well-organized sidechains in the active site and a ligand (Lichtenthaler, 1995). However, the realization that proteins possess intrinsic flexibility emerged with advancements in techniques like NMR spectroscopy, molecular simulations, and advanced imaging methods which allowed for the visualization of protein structures at higher resolutions, capturing subtle changes in atomic positions. As these technologies evolved, discrepancies between static models and experimental data emerged, leading to the recognition that proteins, and more specifically, enzymes, exist in an equilibrium of conformations. In 1958, Koshland’s “induced-fit” model (Koshland, 1958). suggested a more dynamic molecule with interactions that develop in real time as the ligand binds in the active site inducing a conformational change in the enzyme. This model is very similar to the conformational selection model that posits that an enzyme exists in a conformation of substates, and the substrate binds to one of these states thereby stabilizing it in a particular conformation. The difference between these two models is in the order of the conformational change upon ligand binding. The induced-fit model asserts that the ligand binding event predates the conformational change while the conformational selection model asserts that the ligand binding is the result of a conformational change. Regardless of the ordering, both models view enzymes as dynamic entities.

Current thought is that enzymatic activity requires synergy between flexibility and structural stability. Enzyme structure is not just a scaffold to correctly position critical catalytic residues in the active site, it also allows for sampling of functionally productive conformational states (Ramanathan et al., 2011; Ramanathan & Agarwal, 2011). To function well, enzymes must maintain their three-dimensional structure, but be flexible enough to properly bind the substrate, accommodate the transition state and any intermediates, and release the product (Bar-Even et al., 2015; Dellus-Gur et al., 2015). What is known is that these dynamical motions are not random events; they play a pivotal role in enabling enzymes to perform their diverse chemistries. It has also been noted that these dynamic fluctuations span over a range of timescales that depend on the complexity of the motion or the number of atoms involved. The range, from the simplest motion, the vibration of atoms, takes place on the femtosecond timescale; while the movement of side chain methyl groups takes place on the picosecond timescale; the movement of active site residues occurs on the picosecond-nanosecond timescale; and the largest domain movements occur on the microsecond-millisecond timescale (Ojeda-May et al., 2021; Schramm, 2011; Tugarinov et al., 2021; Zeymer et al., 2016).

The Dorothee Kern lab at Brandeis University discovered that the rate of large domain movements taking place on the μs-ms timescale, like the opening and shutting of the lid on adenylate kinase, is the catalytic rate-limiting step. Moreover, in adenylate kinase homologues, differences in their catalytic efficiencies were directly tied to their rates of lid opening and closing (Boehr et al., 2006; K. A. Henzler-Wildman et al., 2007; Wolf-Watz et al., 2004b). How motions on smaller timescales relate to these large

domain movements and in turn enzyme function and structural stability is the topic for ongoing research efforts.

1.6 Modulating Protein Dynamics Through Allostery

In the 1960s, Monod and Jacob coined the term “allosteric” inhibition as a rebuttal to the widely held belief that only steric analogues of the substrate could inhibit enzyme activity. At this time, it was well-documented that many biosynthetic pathways involving a series of enzymes were regulated by the end-product. This allowed for exquisite control in the cell. At the time, it was believed that the enzyme active site was the only place where a substrate could bind and have an impact on the catalytic rate. Therefore, the end-product was regulating enzyme function by competing with the substrate for binding in the active site. Monod and Jacob disagreed. They argued using data on threonine deaminase from Changeux that competitive inhibition in this system was not due to a mutually exclusive binding event, but that there were two distinct binding sites on threonine deaminase. In addition, these two sites were interacting with each other (Monod & Jacob, 1978). Allosteric regulation, or the “communication” between locations distal to each other in geometric space, is an important concept to drug development and signaling. However, an allosteric interaction is not limited to a small molecule-protein binding event, it can be applied in a broader context. For the purposes of this work, the term allostery is defined as any perturbation in a protein that affects the function of a distant site. This broader definition encompasses site-specific perturbations like amino acid substitutions and protein-protein binding events (Ma et al., 1999; Paul & Weikl, 2016).

There is now ample evidence suggesting that changes to locations far from the active site can have a significant impact on the chemistry taking place in the active site (Amor et al., 2016; Boehr et al., 2006; Goldsmith & Tawfik, 2017; Guarnera & Berezovsky, 2016; K. A. Henzler-Wildman et al., 2007; Kamerlin & Warshel, 2010; Leferink et al., 2014; Morley & Kazlauskas, 2005; P. Singh et al., 2015; Warshel & Bora, 2016). Frances Arnold described this phenomenon in her 2018 Nobel lecture. "Mutations happened all over the protein: twenty to thirty angstroms away from the active site where no one could explain them, much less predict them" (F. Arnold, 2019). Unfortunately, Arnold was correct when she stated that a robust method to predict which distal mutations would have an effect on enzyme function does not exist. Currently, the only way to dial into mutations distal to the active site that can fine-tune dynamics and improve catalysis is by random mutagenesis.

1.6 Summary of the Work Presented

The focus of this work was to explore the relationship between enzyme dynamics and substrate specificity. Particularly, elucidating how mutations distal to the active site can influence enzyme catalysis is the focus of chapters 2 and 3. The protein chosen for this work was TEM-1 from the Class A β-lactamase family. β-lactamases are a large family of enzymes that hydrolyze β-lactam antibiotics, conferring resistance to the bacteria producing them. TEM-1 efficiently hydrolyzes antibiotics from the penicillin family but has poor activity against cephalosporins because the added side chain bulk sterically clashes with the active site. In chapter 2, we computationally identified residues that are rigid and allosterically coupled to the active site and we altered the interaction patterns around these important residues. This led to drastic

changes in function and stability. To identify structural alterations that could explain these changes, we crystallized and solved the structures for five TEM-1variants. Chapter 3 is a structural analysis of the models built with data from x-ray crystallographic experiments. In chapter 4, the focus shifts from allosteric contributes to substrate specificity to the active site. This chapter is a discussion on how to modulate substrate specificity by remodeling the active site of firefly luciferase to identify novel luciferase-luciferin pairs for bioluminescent imaging. Specifically, a high-throughput platform was developed to identify target residues for library development. This workflow is broadly applicable to any study of enzyme-ligand binding.

CHAPTER 2

WORK TOWARD MODULATING ENZYME FUNCTION THROUGH ALTERING THE FLEXIBILITY OF RIGID RESIDUES

This chapter is adapted from: "Kolbaba-Kartchner, B.; Kazan, I.C.; Mills, J.H.; Ozkan, S.B. (2021) The Role of Rigid Residues in Modulating TEM-1 β-Lactamase Function and Thermostability. Int. J. Mol. Sci. 22, 2895." I Can Kazan shared first-authorship with Bethany Kolbaba-Kartchner. I Can Kazan conducted all computational work related to DFI, DCI, and MD simulations presented while Bethany Kolbaba-Kartchner performed all RosettaDesign methods and experimental characterization.

2.1 Abstract

The relationship between protein motions (i.e., dynamics) and enzymatic function has begun to be explored in β-lactamases as a way to advance our understanding of these proteins. In a recent study, we analyzed the dynamic profiles of TEM-1 (a ubiquitous class A β-lactamase) and several ancestrally reconstructed homologues. A chief finding of this work was that rigid residues that were allosterically coupled to the active site appeared to have profound effects on enzyme function, even when separated from the active site by many angstroms. In the present work, our aim was to further explore the implications of protein dynamics on β-lactamase function by altering the dynamic profile of TEM-1 using computational protein design methods. The Rosetta software suite was used to mutate amino acids surrounding either rigid residues that are highly coupled to the active site or to flexible residues with no apparent communication with the active site. Experimental characterization of ten designed proteins indicated that alteration of residues surrounding rigid, highly coupled residues, substantially affected both enzymatic

activity and stability; in contrast, native-like activities and stabilities were maintained when flexible, uncoupled residues, were targeted. Our results provide additional insight into the structure-function relationship present in the TEM family of β-lactamases. Furthermore, the integration of computational protein design methods with analyses of protein dynamics represents a general approach that could be used to extend our understanding of the relationship between dynamics and function in other enzyme classes.

2.2 Introduction

Since the 1940s, β-lactam antibiotics, which target a key enzyme in bacterial cell wall biosynthesis, have been the antimicrobial weapon of choice in the war against bacterial infection (Coulson, 1985).The widespread use of β-lactams is likely a consequence of the fact that they are inexpensive to produce and have historically been effective in treating most infections. However, as the use of this class of antibiotics became more widespread, so too did the prevalence of β-lactamase enzymes, which hydrolyze the β-lactam ring and render the antibiotic nonfunctional (Coulson, 1985). Additionally, as new β-lactam antibiotics enter into clinical use, the remarkable adaptivity of β-lactamases complicates efforts to develop novel antibiotics that are resistant to degradation by this class of enzyme (Bush, 2018). The TEM family of β-lactamases has been thoroughly studied to gain insight into the manner in which resistance is achieved (Brandt et al., 2017; Brown et al., 2020; Cortina et al., 2018; Cortina & Kasson, 2018; Gobeil et al., 2019). Despite these efforts, we currently possess an incomplete understanding of the relationship between sequence and function in this enzyme class. A major challenge is that several mutations have been identified that have a significant influence on function, but which

are highly distal from the enzyme active site (M. K. Singh & Dominy, 2012). In addition, even single point mutations (e.g., the well-characterized, M182T substitution), which have minimal effects on enzymatic function can drastically affect the protein's thermostability (Orencia et al., 2001; Wang et al., 2002). Our inability to rationalize the manner in which these thoroughly studied mutations alter enzyme function is suggestive of an incomplete understanding of the sequence-function relationships present in β-lactamases. This in turn limits our ability to develop novel classes of antibiotics that are not substrates for these enzymes (Fair & Tor, 2014).

A possible explanation as to how mutations distal to the active site can still exert influence at a great distance is that they serve to reshape the inherent dynamics of the enzyme(Doucet et al., 2007; Gerek et al., 2009; Gerek & Ozkan, 2011; Kim et al., 2015; Larrimore et al., 2017; Modi et al., 2018; Modi & Banu Ozkan, 2018; Zou et al., 2015). In a recent study, we explored this hypothesis in the TEM-1 β-lactamase using two in silico, dynamics-based metrics: the dynamic flexibility index (dfi) (Gerek & Ozkan, 2011; Kumar, Butler, et al., 2015), which measures the mobility of each residue, and the dynamic coupling index (dci) (Campitelli et al., 2018; Larrimore et al., 2017), which assesses the coupling between distant residues (Zou et al., 2015). Using these two metrics, we characterized TEM-1 and a set of ancestrally reconstructed TEM-1 variants that possess vastly distinct physical properties (i.e., thermostabilities) and functions (i.e., substrate specificity) despite having almost identical conformations (Risso et al., 2013; Salverda et al., 2010; Stiffler et al., 2015; Zou et al., 2015). A major finding of our previous study was that TEM-1 and its ancestral homologues possessed distinct dynamic profiles and that these differences in dynamics appeared to have profound effects on

enzyme function. Namely, rigid residues that are distal from, but highly coupled to, residues in the active site appeared to have substantial effects on protein function (Campitelli et al., 2018; Campitelli, Modi, et al., 2020; Modi & Banu Ozkan, 2018; Stiffler et al., 2015). One intriguing hypothesis that might explain these data is that rigid residues can serve as "hubs" of dynamic communication. This notion has also been validated in the context of disease-causing mutations in other proteins, (Modi et al., 2021) in which mutations to rigid residues that are far from the active site are functionally deleterious (Campitelli, Modi, et al., 2020; Kumar, Glembo, et al., 2015; Modi & Banu Ozkan, 2018; Nevin Gerek et al., 2013).

More recently, we used both dfi and dci to analyze members of the TEM family that either arose in the clinic or were generated via directed evolution (Modi & Banu Ozkan, 2018). In this study, we observed that mutations known to confer resistance to non-native substrates 1) often occur at particularly rigid residues as judged by our dfi metric and 2) appear to allosterically modify the flexibility of catalytic residues within the active site as suggested by our dci metric (Modi & Banu Ozkan, 2018). Collectively, these studies support the hypothesis that rigid residues are of particular importance to the overall dynamics of proteins and may have a substantial impact on protein function if they are allosterically coupled to the active site. If our hypothesis is correct, mutations that alter the identity of allosteric rigid residues (or those in their vicinity) could have substantial effects on enzyme activity; however, the ability to thoroughly explore this hypothesis is challenging. Although extensive datasets comprised of clinically derived TEM family variants (NCBI BioProject Database, 2018, February 1) and additional variants generated via directed evolution (Stiffler et al., 2015) exist, the serendipitous identification of

proteins with multiple mutations in the vicinity of known rigid residues would be unlikely. One potential solution is to use computational protein design methods to specifically target mutations to regions of interest. A major benefit of this approach is the ability to "pre-screen" each combination of mutations *in silico* to exclude variants in which protein folding is not predicted to be energetically favorable.

In this work, computational protein design methods were used to alter the environments surrounding two residues that were identified as being rigid and highly coupled to the active site despite being separated from it by a great distance. Dynamic profiles of each designed protein (hereafter referred to as a "design") were then generated and compared to that of an ancestrally reconstructed variant of TEM-1 (the "Gram-negative common ancestor" or GNCA), which possesses increased thermostability, but reduced activity against ampicillin relative to wild type TEM-1 (Risso et al., 2013). Principal component analysis (PCA) was used to identify designs with dynamic profiles that were predicted to be more similar to GNCA than extant TEM-1, and five designs were characterized in the laboratory. All designs exhibited reduced activity against ampicillin relative to TEM-1, but an increase in thermostability was also observed. Reduced activity against ampicillin and increased thermostability relative to TEM-1 are both features of GNCA. Alternatively, when identical design protocols were applied to flexible residues, that were not coupled to the active site, native-like catalytic abilities and thermostabilities were maintained. Finally, in an effort to further link dynamics to enzyme function, we developed a novel analytical approach termed the "dynamic distance analysis" (dda) that was applied retrospectively to our experimentally characterized proteins. The dda analysis appeared to capture functional differences

between our designed proteins and could be a useful tool for dynamic profile analysis in future studies. Collectively, our results serve to further highlight the importance of allosteric rigid residues in regulating the dynamics of the TEM-1 β-lactamase.

2.3 Materials and Methods

2.3.1 Molecular Dynamics (MD). The AMBER software package was utilized for simulating all β-lactamases in this study. Each system was parameterized with the ff14SB force field and the explicit water model TIP3P (Campitelli, Ozkan, et al., 2020; Salomon-Ferrer et al., 2013). The solvation box was assigned as 16 Å. The system was neutralized by sodium and chloride ions and minimized for 11,000 steps using the steepest descent algorithm. Isothermal, isobaric, and constant number of particles ensemble production trajectories were performed at 300K and 1 bar pressure. For each production, a 1 μs simulation was conducted. The residue covariances were calculated using a 50 ns length window shifted by 10 ns (example: 1–50 ns, 10–60 ns, etc..) over the course of the trajectories.

2.3.2 Dynamic Flexibility Index (dfi). The dfi metric (Gerek & Ozkan, 2011; Kumar, Butler, et al., 2015; Modi & Banu Ozkan, 2018) calculates the relative flexibility/rigidity of a residue in a protein by incorporating the residue covariances. The protein can be modeled with the Elastic Network Model (ENM) in which harmonic springs connect Cαs (Atilgan et al., 2010). Taking the second derivatives of the potential forms a Hessian matrix, *H* Equation (1). The inverse of the Hessian matrix is proportional to the covariance matrix. The models based on ENM cannot capture changes in the dynamics of the designed variants based on Cα positions alone. Therefore, we substituted the inverse of the Hessian with the covariance matrices from MD trajectories to capture

the effect of mutations on the protein conformations. The covariance matrix, *G*, contains the residue covariances, obtained by the MD trajectories Equation (2) and (3) (Bishop, 2006; Kumar, Glembo, et al., 2015; Larrimore et al., 2017; Modi & Banu Ozkan, 2018; Nevin Gerek et al., 2013).

$$[\Delta \mathbf{R}]_{3Nx1} = [\boldsymbol{H}]^{-1}_{3Nx3N}[\mathbf{F}]_{3Nx1} \tag{1}$$

$$[\Delta \mathbf{R}]_{3Nx1} = [\boldsymbol{G}]_{3Nx3N}[\mathbf{F}]_{3Nx1} \tag{2}$$

$$df i_i \quad = \quad \frac{\sum_{j=1}^{N} \left|\Delta R^j\right|_i}{\sum_{i=1}^{N}\sum_{j=1}^{N} |\Delta R^j|_i} \tag{3}$$

The residue response vector (ΔR) is the resultant vector containing the fluctuation responses from multiplying the covariance matrix with the force vector, F. $\left|\Delta \boldsymbol{R}^j\right|_i$ denotes the magnitude of the residue response fluctuation vector of position *i*, when *j* is exposed to a random force vector.

2.3.3 Dynamic Coupling Index (dci). The dynamic coupling index (dci) (Bishop, 2006; Larrimore et al., 2017; Modi & Banu Ozkan, 2018) measures the degree of dynamic coupling between two residues. Namely, it captures the strength of displacement of a residue *i* upon perturbation of a distinct residue *j*, relative to the average fluctuation response of position *i* when all the positions within a structure are perturbed. Generally, this metric is used to establish the communication between a functionally important residue and other residues within the protein that are many angstroms away. The dynamic coupling index of a given residue *i* is calculated using the equation below Equation (4):

$$dci_i = \frac{\sum_j^{N_{Functional}} \left|\Delta R^j\right|_i / N_{Functional}}{\sum_{j=1}^{N} |\Delta R^j|_i / N} \quad (4)$$

where $\left|\Delta \boldsymbol{R}^j\right|_i$ corresponds to the magnitude of the residue response vector (ΔR) for residue i when residue j is perturbed. The dci score thus provides information on the allosteric behavior of a location associated with active site dynamics. A high dci value implies strong coupling between active sites, inversely, a low scoring position is regarded as weakly coupled to the active site (Bishop, 2006; Larrimore et al., 2017; Modi & Banu Ozkan, 2018).

2.3.4 Dynamic Distance Calculation. Principal Component Analysis (PCA) was used to compare and cluster the flexibility profiles of the designed TEM-1 variants with respect to TEM-1 and GNCA. However, because the output of a PCA is dependent on the input data, the calculated distances between any designed protein and TEM-1 or GNCA can change with the inclusion of new or distinct data points (e.g., a different set of designed proteins). To account for this, we employed an iterative, random sampling approach to capture the relative distance of a designed protein from TEM-1 and from GNCA (Figure Figure 2.3.4.1).

For every designed TEM-1 variant, a dataset containing the target design, TEM-1, GNCA and an additional seven randomly chosen designs was constructed and used to generate a PCA. Namely, the dfi profiles of these ten proteins were merged into a matrix, X, of dimension Equation (5)

$$(m \times n) \quad (5)$$

Here, *m* is the total number of datasets that are clustered together, which each have *n* number of attributes (*n* = total number of residues). Singular value decomposition of *X* was then carried out as follows Equation (6):

$$[X]_{m\times n} = [U]_{m\times m}[\Sigma]_{m\times n}[V]_{n\times n} \quad (6)$$

Here, *U* and *V* are unitary matrices with orthonormal columns and are called left singular vectors and right singular vectors, respectively, and Σ is a diagonal matrix with diagonal elements known as singular values of *X*.

The singular values of *X*, by convention, were arranged in a decreasing order of their magnitude, σ = {σi} representing the variances in the corresponding left and right singular vectors. The set of the highest singular values (representing the largest variance in the orthonormal singular vectors) can be interpreted to show the characteristics in the data *X* and the right singular vectors create orthonormal basis which spans the vector space representing the data. The left singular vectors contain weights indicating the significance of each attribute in the dataset as Equation (7):

$$w_i = \sum_{k=1}^{r} \sigma_k |u_{ik}| \quad (7)$$

Using these features of the decomposed singular vectors, we created another matrix, *X** using only the highest three singular values which mimics the basic characteristics of the original dataset. It can be represented as Equation (8):

$$[X^*]_{mxr} = [\boldsymbol{V}^*]_{mxr}[\Sigma^*]_{rxr} \quad (8)$$

Here, Σ* contains only the largest 3 singular values and V^* contains the corresponding right singular vectors. The data were then clustered hierarchically based on the pairwise distance between different proteins in the reconstructed dfi data with reduced dimensions. The distance between designed protein, j_1, and TEM-1, j_2, was computed in the reduced dimension using three principal components Equation (9):

$$d_{12} = \sqrt{\sum_{i=1}^{3} (X_i^{*j_1} - X_i^{*j_2})^2} \tag{9}$$

We also calculated the distance between each designed TEM-1 variant and GNCA to measure the similarity in their flexibility profiles. The random selection of dataset was repeated a thousand times to create a diverse distance distribution and we called this distance profile analysis dynamic distance analysis (dda). The distributions were fit to a Gaussian mixture model with a Dirichlet prior to estimate the density and the mean of the dynamic distances (Gibson et al., 2009). The distributions and the mean distances were utilized for selecting the designed proteins that cluster close to GNCA and far from TEM-1 (Figure 2.3.4.1).

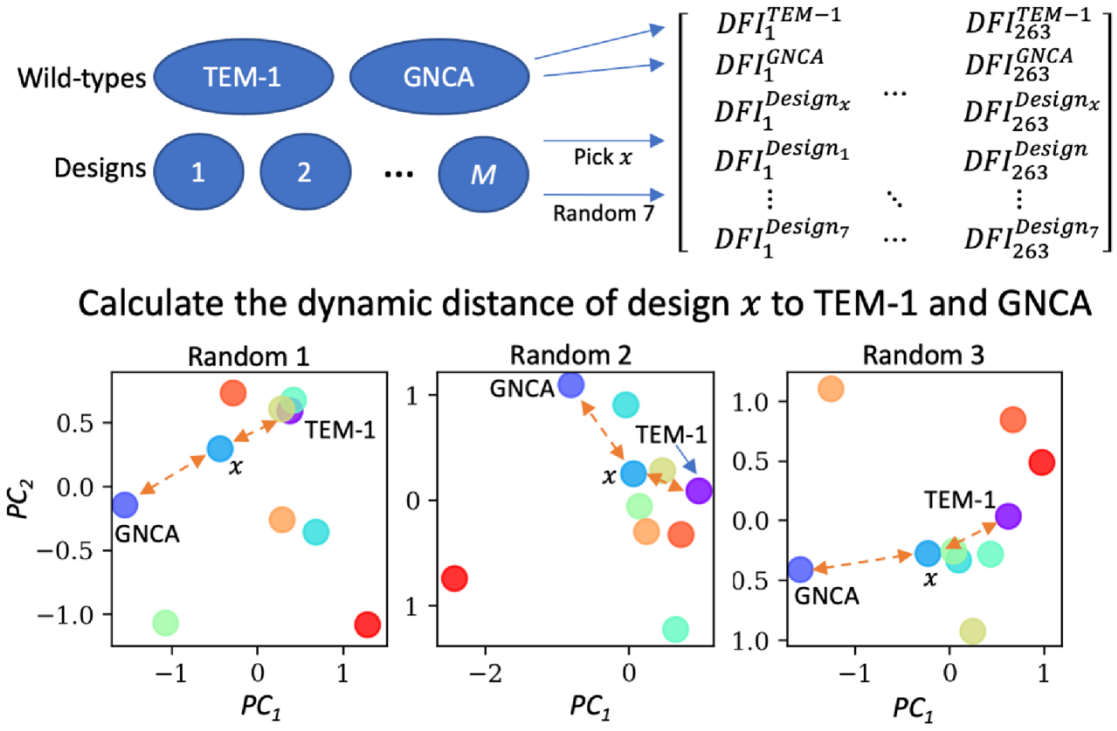


Figure 2.3.4.1. Schematic of the dynamic distance calculation process. The dynamic profile of each design (using the dfi metric) is clustered using PCA in a set composed of TEM-1, GNCA, and seven randomly chosen designs. The dynamic distance of the design from TEM-1 and GNCA is calculated. Notably, the dynamic distance of the designed protein from TEM-1 and GNCA varies according to the set of proteins incorporated. To capture a statistically accurate distribution, this procedure is iterated a thousand times, each time varying the set of designed proteins.

2.3.5 Rosetta Design Protocol. A high-resolution (1.8 Å) structure of TEM-1 (PDB ID: 1btl) was processed to remove waters and non-proteinogenic molecules. The resulting structure was subjected to an energy minimization using the Rosetta relax protocol (Conway et al., 2014a; Nivón et al., 2013). (Detailed descriptions of all computational protocols used in this study can be found in Appendix B). The relaxed 1btl structure was used as an input to the DesignAround protocol within Rosetta using the ref15 score function (Alford et al., 2017a). This algorithm first identifies spheres with user-defined

radii around a defined residue. Residues within these "design spheres" were subjected to *in silico* mutagenesis, conformational sampling and backbone minimization.

2.3.6 Protein Expression and Purification. A pET24b plasmid encoding the gene for GNCA was a generous gift from Professor Jose Sanchez-Ruiz (Universidad de Granada). Genes encoding rigid design variants were codon-optimized for expression in *Escherichia coli* cells. The native TEM-1 N-terminal periplasmic localization signal peptide (MSIQHFRVALIPFFAAFCLPVFA) was appended to the beginning of each gene. To facilitate purification, a C-terminal 6xHis affinity tag was added to the end of each gene. Genes encoding each rigid design were synthesized by IDT (Coralville, IA). The gene for wildtype TEM-1 was amplified from a pET21b vector using PCR. Genes encoding the rigid designs and TEM-1 were subcloned into the pET29b vector using the Gibson Assembly (Gibson et al., 2009) at a site that placed them under the control of the T7lac promoter. Genes encoding the uncoupled flexible residue variants were synthesized and cloned into pET29b vectors by GenScript (Piscataway, NJ).

The sequences of all plasmids containing TEM-1, GNCA, rigid or flexible designs were confirmed by Sanger sequencing and were transformed via electroporation into BL21 Star (DE3) *E. coli* cells. Cells containing plasmids encoding GNCA were grown in lysogeny broth (LB) at 37 °C with shaking at 250 rpm until an $O.D._{600}$ of ~0.8 was reached. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 1 mM to induce expression; cells were grown for 3 h post induction. Cells containing plasmids encoding TEM-1 were grown in LB media at 20 °C with shaking at 220 rpm until an $O.D._{600}$ of ~0.8 was reached. Induction was again carried out with 1 mM IPTG and was allowed to proceed for 8–12 h. Cells containing plasmids

encoding the rigid and flexible design variants were grown in 2xYT media to confluence overnight, and pelleted by centrifugation. After resuspension in fresh 2xYT media, protein expression was induced with 1 mM IPTG and cells were grown for an additional 20 h at 20 °C with shaking at 220 rpm.

After expression, the cells were pelleted via centrifugation at 4,100x *g* for 15 min and the media was discarded. The cells were resuspended in TBS (50 mM Tris pH 8.0, 500 mM NaCl) and were again centrifuged at 4,100x *g* for 15 min; the supernatant was discarded. The pellet was incubated at room temperature for 15 min with SET buffer (20% sucrose, 1 mM ethylenediaminetetraacetic acid (EDTA), 30 mM Tris pH 8.0, 1 mg/mL lysozyme). After centrifugation at 4,100x *g* for 15 min, the supernatant was decanted and saved. The cells were then shocked to release the periplasmic contents with ice cold 100 mM MgCl2 at a 1:15 ratio of cell pellet weight to solution volume. Cells were vigorously agitated on ice for 15-30 min then centrifuged with the saved soluble fraction from the first stage at 4 °C for 60 min at 12,000x *g*.

The supernatant was then loaded onto a 5 mL nitrilotriacetic acid agarose (Ni-NTA) (Millipore Sigma, Burlington, MA) column, washed with 5 column volumes of a low imidazole buffer (25 mM Tris pH 8.0, 150 mM NaCl, 15 mM imidazole), and eluted with a high imidazole buffer (25 mM Tris pH 8.0, 150 mM NaCl, 500 mM imidazole). All proteins were then subjected to a second purification step using anion exchange chromatography: Proteins were concentrated to a volume of 0.5–1 mL, diluted into the loading buffer (50 mM Tris, pH 9.0, 50 mM NaCl) and loaded directly onto the 5 mL Hi Trap Q Fast Flow column (Millipore Sigma, Burlington, MA). The column was washed

with 5 column volumes of the loading buffer and eluted with 50 mM Tris, pH 9.0 250 mM NaCl. Protein purity was verified by SDS-PAGE (Figure 2.4.4.1).

2.3.7 Circular Dichroism Characterization of Protein Folding and Stability. Far-ultraviolet circular dichroism (CD) measurements were performed in triplicate on a Jasco J-815 spectrophotometer (Jasco, Inc, Easton, MD) equipped with a Peltier temperature controller. Wavelength scans were measured from 300 to 180 nm at room temperature with 1 nm steps using a 1 nm bandwidth, 5 nm/min scan rate; reported data represent an average of three independent scans. Thermal melts were monitored by the absorption signal at 222 nm with a temperature slope of 5 °C/min. For wavelength scans and thermal melts, the purified protein was in a TBS buffer (10mM Tris 50 mM NaCl, pH 7.0) in a cuvette with a 1 mm pathlength. Protein concentrations were calculated in triplicate using the absorbance at 280 nm and absorption coefficients as calculated by the ProtParam tool in the Expasy software suite (Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., 2005). Protein concentrations ranged between 0.18–0.25 mg/mL for all scans. Thermal melt curves were fitted using nonlinear regression least squares fit with the Hill equation in the GraphPad Prism version 9.0.0 for Windows, GraphPad software, San Diego, California, USA.

2.3.8 Minimal Inhibitory Assays. Minimal inhibitory concentrations of ampicillin (MIC_{amp}) were performed in triplicate on 96-well plates (Wiegand et al., 2008). For each designed protein, TEM-1 and GNCA, five colonies were picked from a fresh agar plate and used to inoculate a 5 mL culture of LB, which was grown to confluence overnight at 37 °C. Overnight cultures were diluted in LB with 1 mM IPTG to a final working concentration of 5 × 105 cfu/mL. Three stock solutions of ampicillin were independently

prepared at 6000 μg/mL in LB with 1 mM IPTG and each solution was subsequently diluted in steps of 0.5 through the addition of LB with 1 mM IPTG to yield a final range of concentrations of 6-3000 μg/mL. The ampicillin concentrations for GNCA and the rigid designs were prepared at 400 μg/mL in LB with 1 mM IPTG and each solution was diluted in steps of 0.6 for a final concentration range of 2-200 μg/mL. The 96-well plates were covered with a fitted lid and incubated at 37 °C for 20 h. All optical density measurements were carried out at 600 nm using a SpectraMax M5 (Molecular Devices, LLC, San Jose, CA); the absorbance of the buffer was subtracted from each measurement. To establish the lowest concentration of antibiotic that inhibited growth, a buffer-subtracted value > 0.1 was used as the threshold for bacterial growth in each well. The MIC_{amp} was determined to be the lowest concentration of ampicillin that inhibited growth of the *E. coli* cells.

2.4 Results and Discussion

2.4.1 Computational Analysis Using dfi and dci. Our efforts to better understand the relationship between protein dynamics and function began by identifying a TEM-1 variant that could serve as a basis of comparison to the wild type protein. Recently, the putative sequences of ancestral TEM-1 were predicted using Bayesian bioinformatics (Risso et al., 2013). Three ancestral TEM family homologues (the Gram-negative and Gram-positive common ancestor, PNCA; the Gram-negative common ancestor, GNCA, and enterobacteria common ancestor, ENCA) were observed to possess distinct physical and biochemical properties when characterized in the laboratory (Risso et al., 2013). This is likely a consequence of the fact that these proteins are thought to have existed at different times in the evolutionary history of this enzyme (Risso et al.,

2013). We chose to focus our efforts on the ancestral homologue GNCA because its properties differ more substantially from TEM-1 than the other variants. Despite sharing > 50% identical residues (Figure 2.4.1.1.A), nearly identical folds (1.3 Å root-mean-square deviation (RMSD) over all Cαs, Figure 2.4.1.1.B), and conserved catalytic residues (Figure 2.4.1.1.C), GNCA unfolds at a temperature that is ~35 °C higher than wild type TEM-1.

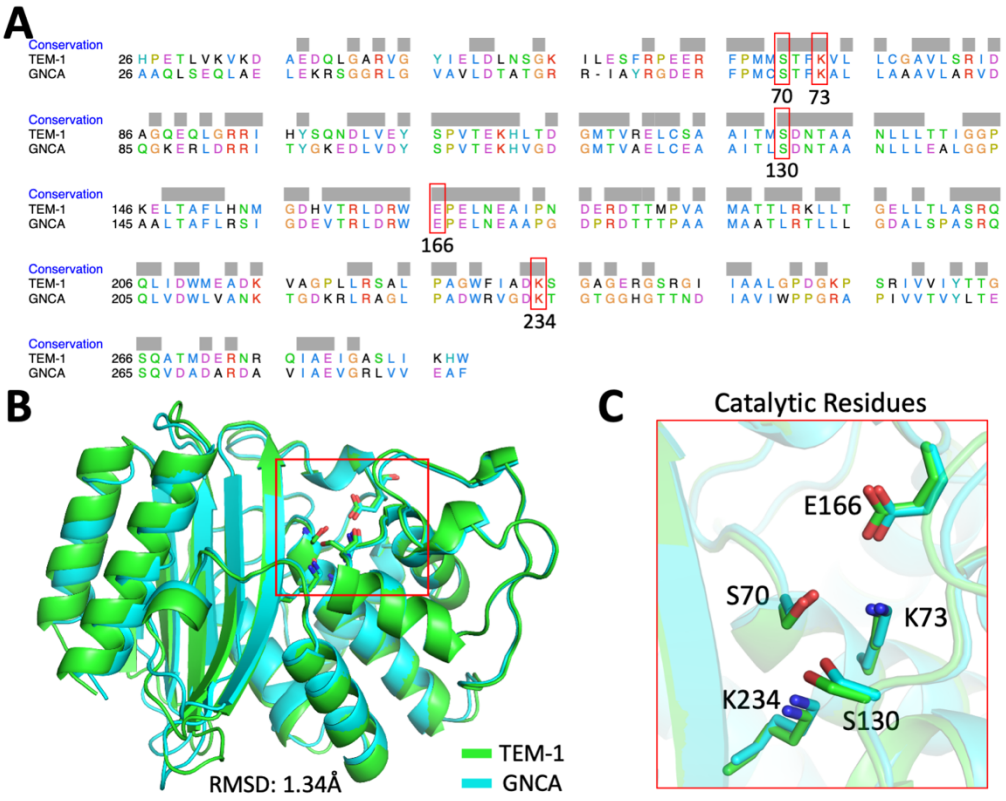


Figure 2.4.1.1 Differences in sequence and structure between TEM-1 and its ancestral variant GNCA. (A) Sequence alignment (Ambler numbering) (Ambler et al., 1991) of TEM-1 and GNCA shows a 54% sequence identity; conserved active site residues are highlighted in red boxes. (B) The crystal structures of TEM-1 (PDB ID: 1btl, green (Jelsch, C.; Mourey, L.; Masson, J.M.; Samama, 1993)) and GNCA (PDB ID: 4b88, cyan (Risso et al., 2014)) are superimposed and the catalytic residues are shown as sticks within a red box. The low root-mean-square deviation (RMSD) indicates a high conservation of structure. (C) Active site residues in TEM-1 and GNCA are shown in green and blue sticks, respectively.

Furthermore, GNCA appears to be a "substrate generalist" in that it possesses measurable (but reduced) activity against penam antibiotics (e.g., penicillin and ampicillin) relative to TEM-1, while simultaneously possessing a far greater ability to degrade the bulkier cepham antibiotics (e.g., cefotaxime) (Figure 2.4.1.2) (Risso et al., 2013).

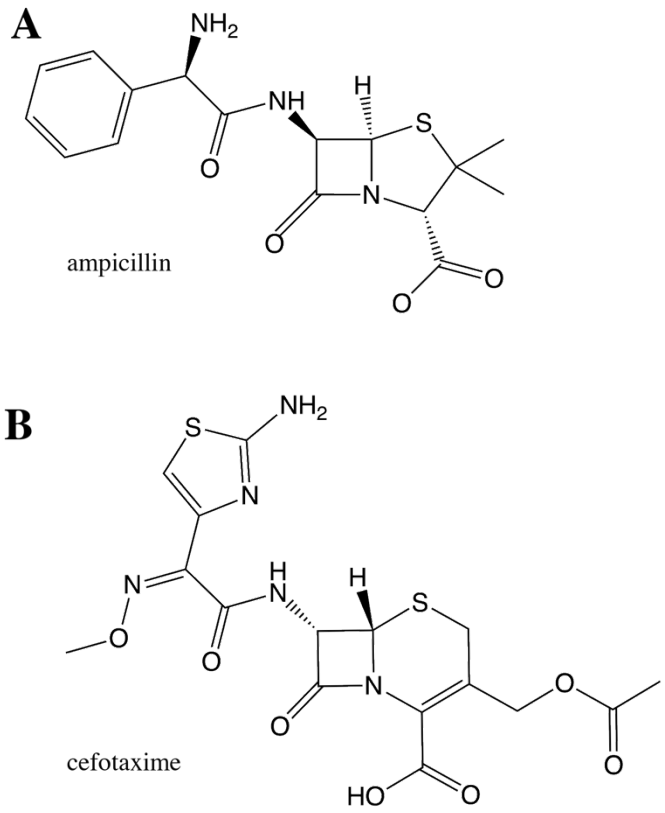


Figure 2.4.1.2 Chemical structure of ampicillin and cefotaxime. (A) Ampicillin, a member of the penam, or penicillin family antibiotics. (B) Cefotaxime, a member of the cepham, or the third generation cephalosporin family antibiotics.

It is difficult to rationalize the substantial differences in function and stabilities that are observed in GNCA and TEM-1 in light of the high sequence identity and structural similarities that exist for these proteins. Previous studies in our laboratory (Modi & Banu Ozkan, 2018; Zou et al., 2015) suggested that the inherent dynamics of both TEM-1 and GNCA might play a role in regulating their functions. To further explore this, we

analyzed the dynamic profiles of both proteins using two metrics developed in our group: The Dynamic Flexibility Index (dfi) and the Dynamic Coupling Index (dci). The dfi method (Butler et al., 2015; Kumar, Glembo, et al., 2015; Nevin Gerek et al., 2013) is based on Linear Response Theory and Perturbation Response Scanning (Atilgan et al., 2010) and calculates the resilience of a given residue to random force perturbations applied to other residues in the protein. A given amino acid's dfi value is therefore related to the relative conformational entropy (i.e., flexibility) of that residue with respect to the rest of the protein. A residue with a high dfi value indicates high flexibility; conversely, a low dfi value indicates rigidity. The dci metric (Larrimore et al., 2017; Modi & Banu Ozkan, 2018) is derived from the same theoretical origin as dfi and is used to quantify the degree to which two residues are dynamically coupled in terms of correlated motions. A high dci value between a pair of residues that do not interact directly indicates allosteric coupling and suggests that a perturbation to one residue will be transmitted to the other even over long distances. A low dci score implies a weak coupling between a residue pair, and no strong communication channel between them is expected.

When we applied the dfi and dci analyses to extant TEM-1 and a set of reconstructed ancestral homologues including GNCA (Modi & Banu Ozkan, 2018; Zou et al., 2015), our analyses indicated that rigid residues (i.e., those with low dfi scores) that are highly coupled to the active site can contribute substantially to protein function. In this study, we hoped to further explore the importance of rigid residues to protein function by altering the identity of amino acids in their vicinity.

We selected two residues in TEM-1 (V44 and V262) as targets for our study. Not only do both residues have low dfi scores (%dfi value < 0.2) (Figure 2.4.1.3.A), but they

are highly coupled to the active site (%dci > 0.7) (Figure 2.4.1.3.B). These two residues were of particular interest to us because they are over 10 Å away from the active site and are located on adjacent β-strands with side chains facing opposite domains. We also identified three distal, flexible residues in TEM-1 (K55, P226, and K256) with high dfi scores (%dfi > 0.8) (Figure 2.4.1.3.A) and low coupling to active site residues as evaluated by the dci metric (%dci < 0.4) (Figure 2.4.1.3.B) and over 10 Å away from the active site to serve as controls. Alteration of the protein environments surrounding allosteric rigid residues would be expected to substantially modify protein function if our hypothesis is correct. Alternatively, modification of amino acids surrounding flexible residues with low dynamic coupling to the active site would be expected to result in proteins with native-like functions. All the allosteric rigid and uncoupled flexible residues we targeted for design are over 10 Å from the nearest catalytic residue, which suggests that mutations in their vicinities should only have an indirect effect on the active site unless other factors (e.g., dynamic coupling) are at play.

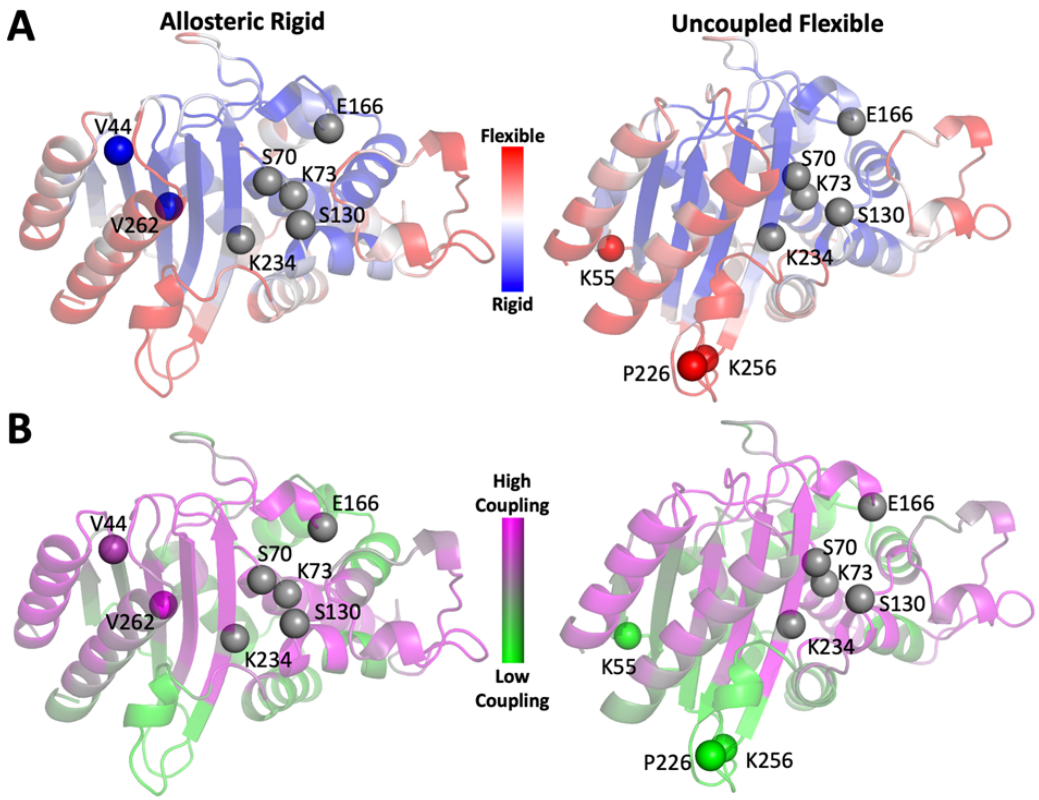


Figure 2.4.1.3. The dfi (A) and dci (B) values of each residue in TEM-1 are calculated and mapped onto the structure of TEM-1, which is shown as color coded cartoons. Catalytic residues are shown as grey spheres. Rigid and flexible residues used in this study are shown as spheres that are colored by either their dfi (A) or dci (B) score. Allosteric rigid residues, V44 and V262, have low dfi scores and high allosteric dynamic coupling with the active site residues. Residues K55, P226, and K256 are both highly flexible and exhibit low allosteric dynamic coupling to the active site.

2.4.2 Computational Design of TEM-1 Variants. To alter the amino acid compositions surrounding both the rigid and flexible residue positions, we used the Rosetta computational protein design suite (Leaver-Fay, Tyka, et al., 2011). The Rosetta software employs a Monte Carlo sampling protocol to randomize the identity and conformation (rotamer) of a randomly chosen residue; the fitness of the mutated protein is then assessed using the Rosetta energy function (Alford et al., 2017). In the course of a

single design trajectory, the Monte Carlo sampling algorithm is applied iteratively to a set of user-defined residues (see below).

We sought to develop a computational protocol within Rosetta that would substantially alter the chemical properties of the native amino acids without negatively affecting the protein's ability to fold. To do this, the RosettaDesign algorithm (Kuhlman et al., 2003) was used to randomly mutate residues within "design spheres" that had radii from 8-12 Å surrounding each of the target residues (Figure 3A). Slight alterations to the conformation of the peptide backbone were allowed only for residues that fell within the design sphere. A second shell was also defined that extended 4 Å beyond the inner design sphere. Residues in this shell were precluded from mutating but were energetically minimized in the context of adjacent, mutated residues. Independent design trajectories were carried out for all rigid and flexible residues. The two rigid (V44 and V262) and three flexible (K55, P226 and K256) residues that served as targets for our studies were also prohibited from mutating during the design calculations (Figure 2.4.2.1). Finally, catalytic residues (S70, K73, S130, E166, K234) were also maintained as their native identities and conformations during the design process. The designed proteins contained between two and eleven mutations with an average of seven mutations per protein. Ultimately, 64 unique designed proteins were generated using this approach.

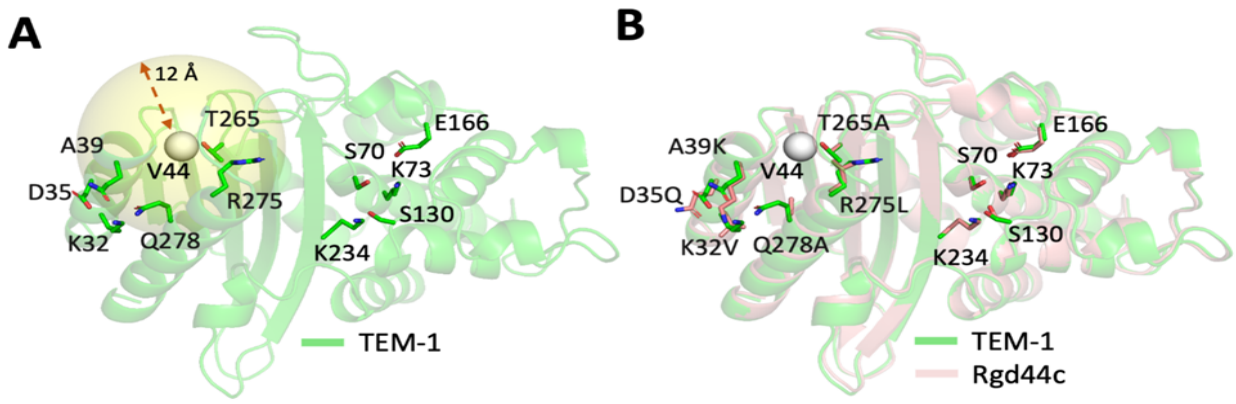


Figure 2.4.2.1 Our general computational protein design strategy is shown schematically using the designed protein Rgd44c as an example. (A) Residues within an 8–12 Å sphere surrounding a given residue (V44 in this example) are considered as candidates for mutation. (B) A combination of mutations surrounding the target residue are generated using the RosettaDesign algorithm and scored using the Rosetta energy function. An overlay of the Rgd44c design model with TEM-1 (B) indicates that this design protocol creates a diversity of mutations within the design sphere while leaving active site residues untouched. The target rigid residue (V44) is shown as a white sphere in both panels. Both catalytic and designed residues are shown as sticks.

2.4.3 Selection of the Designed Proteins Using Flexibility Profiles. To assess how the computationally designed mutations affected TEM-1 dynamics, we subjected all designed proteins to a 1 μs molecular dynamics (MD) simulation followed by analysis using the dfi metric (Figure 2.4.3.1.A). In order to rapidly compare the dfi profiles of our designed proteins to those of TEM-1 and GNCA, we used a 2D principal component analysis (PCA). The PCAs both simplified our data and allowed for the facile visualization of relationships between the calculated dynamic profiles of the designed proteins (Figure 2.4.3.1.B). PCAs generated from our rigid designs showed a diverse distribution in both the first and second principal components (Figure 2.4.3.1.C). On the PCA, several designed proteins were positioned relatively closer to GNCA in both components. We chose a subset of five such designs in which the allosteric rigid residues

had been targeted (henceforth referred to as “rigid designs”) for experimental characterization (Figure 2.4.3.1.C).

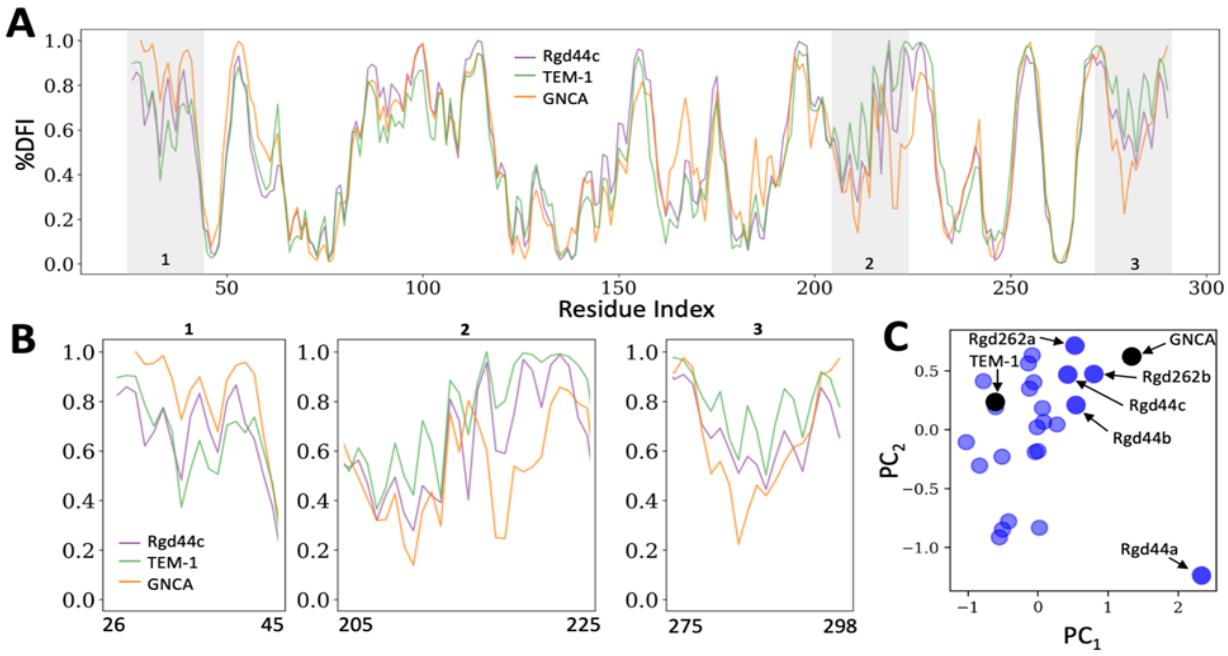


Figure 2.4.3.1 Dynamic analyses of TEM-1, GNCA, and the rigid design, Rdg44c. (A) Depiction of the dfi profile of TEM-1 (green), GNCA (orange) and variant Rgd44c (purple); Rdg44c is chosen as an example for illustrative purposes. (B) Portions of the full dfi profile of each protein (A) are expanded to highlight dynamic differences between the three proteins. A shift towards a GNCA-like dfi profile is an indication of a change in dynamical characteristics of a protein. (C) Principal Component Analysis (PCA) of the rigid designs. The first (x-axis) and second (y-axis) principal components have weights of 3.5 and 2.7, respectively. Designs chosen for experimental characterization are highlighted using darker colors and labeled with the design name.

Four of the five rigid designs (Rdg44b, Rdg44c, Rdg262a, and Rdg262b, where the number in each name corresponds to the rigid residue that was targeted in the design calculations) clustered slightly away from TEM-1 and towards GNCA on both axes of the PCA; alternatively, Rdg44a, clustered near GNCA on the first principal axis but appeared as an outlier on the second axis. We hoped that experimental characterization of Rdg44a might help elucidate the parameters captured in each of the two principal components. It should be mentioned that only four among the five rigid designs that were

chosen for characterization had Rosetta scores that were lower (lower Rosetta scores imply lower energies) than TEM-1. The Rosetta score of Rdg262a was higher than TEM-1, but we selected this design for experimental characterization due to the fact that it clustered near GNCA in both axes of the PCA.

To analyze the designed proteins in which flexible, uncoupled residues were targeted (henceforth referred to as "flexible designs"), we generated a PCA in which all flexible design candidates were compared to TEM-1, GNCA and all the rigid designs including those that were not selected for characterization (Figure 2.4.3.2). Although a wide distribution of flexible designs was observed in this PCA, many of them clustered near TEM-1; a smaller subset clustered near the rigid designs we previously selected for characterization. To avoid biases that might have arisen if we chose only flexible designs that clustered with TEM-1 for analysis, we opted to experimentally characterize four flexible designs (Flx226a, Flx226b, Flx226c and Flx55) that clustered near the rigid designs chosen for experimental characterization and only one (Flx256) that clustered near TEM-1 (Figure 2.4.3.2).

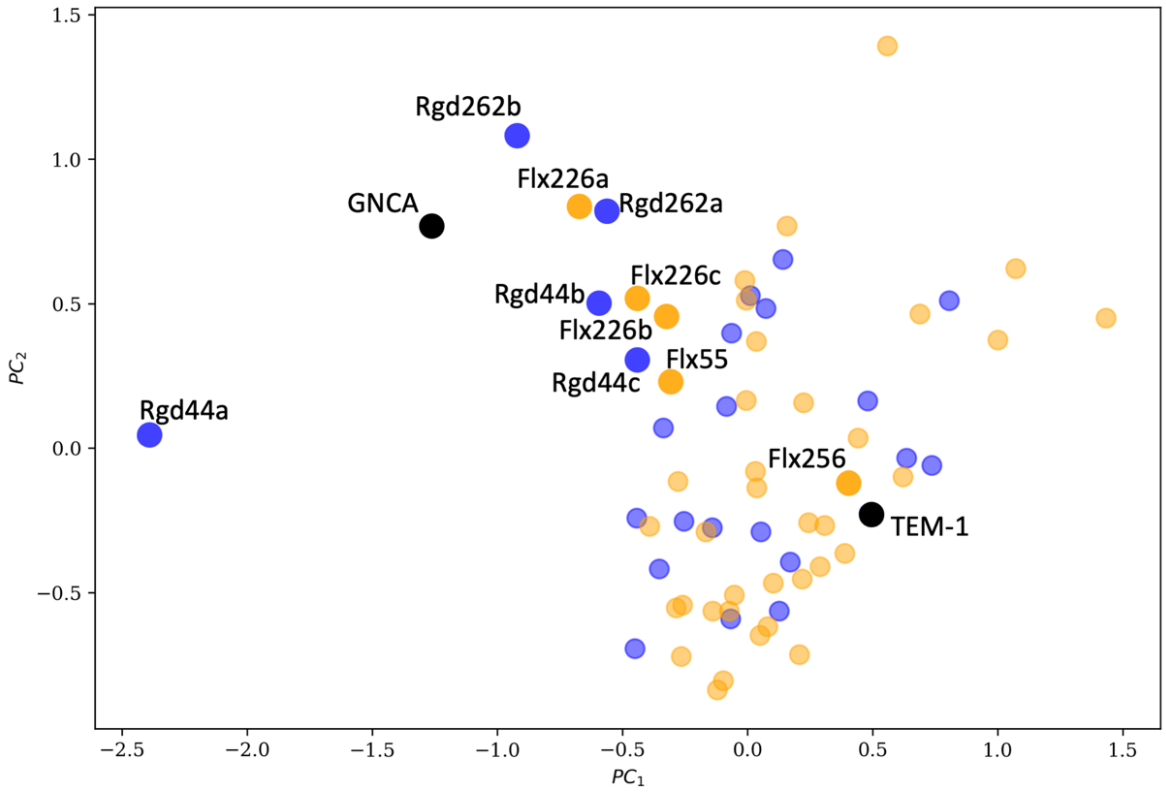


Figure 2.4.3.2 PCA of a selection of the flexible and rigid designed proteins. The rigid designs with allosteric dynamic coupling to the active site are marked with blue dots. Uncoupled flexible designs are marked with orange dots. TEM-1 and GNCA are shown as black dots. For both rigid and flexible designs, the variants chosen for experimental characterization are named and highlighted with darker colors.

Although clustering in similar locations in the PCA would suggest that the two proteins should have similar properties, it is difficult to infer what feature is represented on each axis of the PCA. We hoped that the diverse selection of proteins chosen for characterization would therefore provide information regarding whether rigid residues serve as hubs of dynamic control and also whether or not the PCA is a useful metric for discriminating between proteins with different activity and thermostabilities.

2.4.4 Experimental Analysis of the Designed Proteins. As GNCA and TEM-1 differ substantially with respect to thermostability (90.3 °C and 56.4 °C, respectively) and

activity against penam β-lactam antibiotics (GNCA is ~2 orders of magnitude less efficient at degrading ampicillin than TEM-1), we chose to focus our analyses of the designed proteins on these characteristics. To do this, genes encoding each of the selected rigid and flexible designs were first cloned into the pET29b expression plasmid. Sequenced confirmed plasmids were transformed into a BL21 Star (DE3) *E. coli* expression strain in preparation for further analyses.

We assessed the resistance of our designed proteins to penam β-lactams by establishing the minimal inhibitory concentration of ampicillin (MIC_{amp}) for each of our designed proteins using the protocol of Wiegand et al. (Wiegand et al., 2008). (See Materials and Methods for detailed protocols). Briefly, BL21 Star (DE3) cells harboring a pET29b plasmid that contained a gene encoding one of our variants were grown in a liquid medium containing a range of ampicillin concentrations and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), which induced overexpression of our TEM-1 variants. The ability of cells to grow at each ampicillin concentration was determined by measuring the optical density at 600 nm ($O.D._{600}$); the lowest antibiotic concentration that inhibited growth was recorded. All rigid designs were observed to exhibit either minimal or no activity against ampicillin (Table 1). The two rigid designs that showed the highest activity against ampicillin, Rdg44c and Rdg262b, had MIC_{amp} values of 26 μg/mL, which is two orders of magnitude less efficient than wild type TEM-1 (MIC_{amp} = 1500 μg/mL), but is only half that of GNCA (MIC_{amp} = 43 μg/mL). Alternatively, the MIC_{amp} values of all the flexible designs were in the range of 375-1500 μg/mL (Table 1) which is on par with wild type TEM-1.

Table 1. Minimal Inhibitory Concentrations (MIC_{amp}) and melting temperatures of the TEM-1 variants.

Variant	Minimal Inhibitory Concentration of ampicillin MIC_{amp} (μg/mL)	Melting Temperature T_m (°C)
GNCA	43	90.3
TEM-1	1500	56.4
Rdg44a	< 2**	NM
Rdg44b	< 2	63.1
Rdg44c	26	66.4
Rdg262a	< 2**	NM
Rdg262b	26	56.4
Flx226a	1500	57.4
Flx226b	375	53.2
Flx226c	1500	55.6
Flx256	750	58.1
Flx55	750	58.5

Minimal Inhibitory Concentrations for ampicillin (MIC_{amp}) values were determined in lysogeny broth. Melting temperatures (T_m) were determined using circular dichroism. NM indicates that a T_m was not established for this protein due to aggregation during purification. **Because these variants precipitated out of solution during purification, it is difficult to know whether these values accurately reflect their activities *in cellulo*.

Two possible explanations for the lack of activity against ampicillin observed in our rigid designs are: 1) that only poor protein expression was achieved or 2) that they did not fold into native-like structures; neither of these possibilities are directly examined in MIC assays. We therefore expressed and purified each of the designed proteins and assessed their abilities to adopt native-like structures using circular dichroism (CD) spectroscopy. All designed proteins were observed to express solubly (Figure 2.4.4.1).

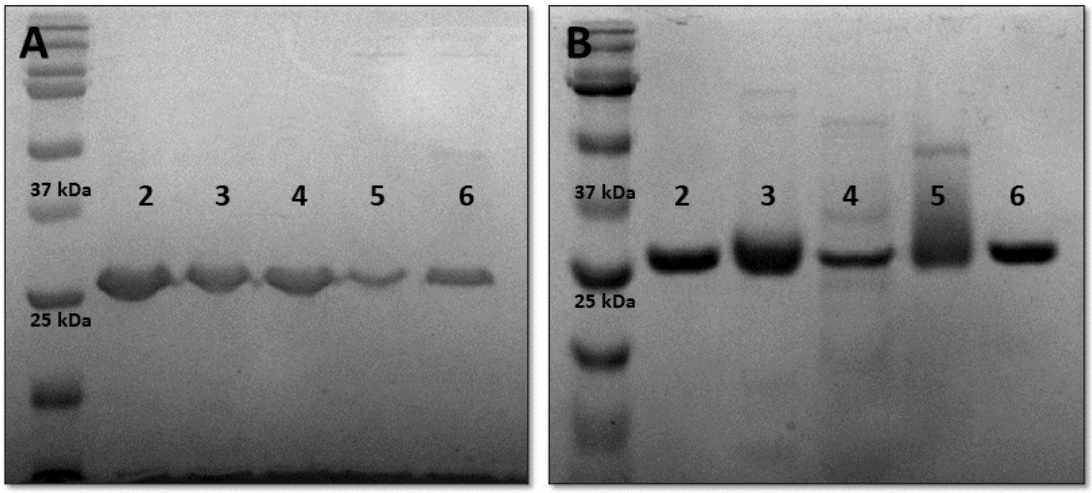


Figure 2.4.4.1. 12% SDS PAGE gels of the purified designed proteins. The gels were stained with Coomassie Brilliant Blue G-250. For the gels, proteins were heat denatured. The protein standard (lane 1) is Bio-Rad Precision Plus Protein Kaleidoscope Prestained Protein Standards (A) Flx226a (lane 2) Flx226b (lane 3) Flx226c (lane 4) Flx256 (lane 5) Flx55 (lane 6) (B) TEM-1 (lane 2) GNCA (lane 3) Rdg44b (lane 4) Rdg44c (lane 5) Rdg262b (lane 6).

However, two of the rigid designs, Rdg44a and Rdg262a, had a high propensity to aggregate during the purification process, which precluded further characterization. In contrast, no aggregation of any of the flexible designed proteins was observed throughout the purification process. We subjected all purified proteins to both wavelength scans and thermal melts using CD (see Materials and Methods), which allowed determination of the

melting temperature (T_m) of each protein (Figure 2.4.4.2). The T_ms of all flexible designs fell into a range (53.2 °C to 58.5 °C) that was within ~3 °C of the T_m of TEM-1 (56.4 °C, Table 1). Alternatively, the T_ms of the rigid designs varied greatly. Although the least stable of the allosteric rigid designs (Rdg262b) exhibited a Tm that was on par with TEM-1, two others exhibited marked increases in stability. Namely, Rdg44b and Rdg44c were measured to have T_ms of 63.1 °C and 66.4 °C, respectively, which correspond to increases of ~6 °C and 10 °C relative to TEM-1.

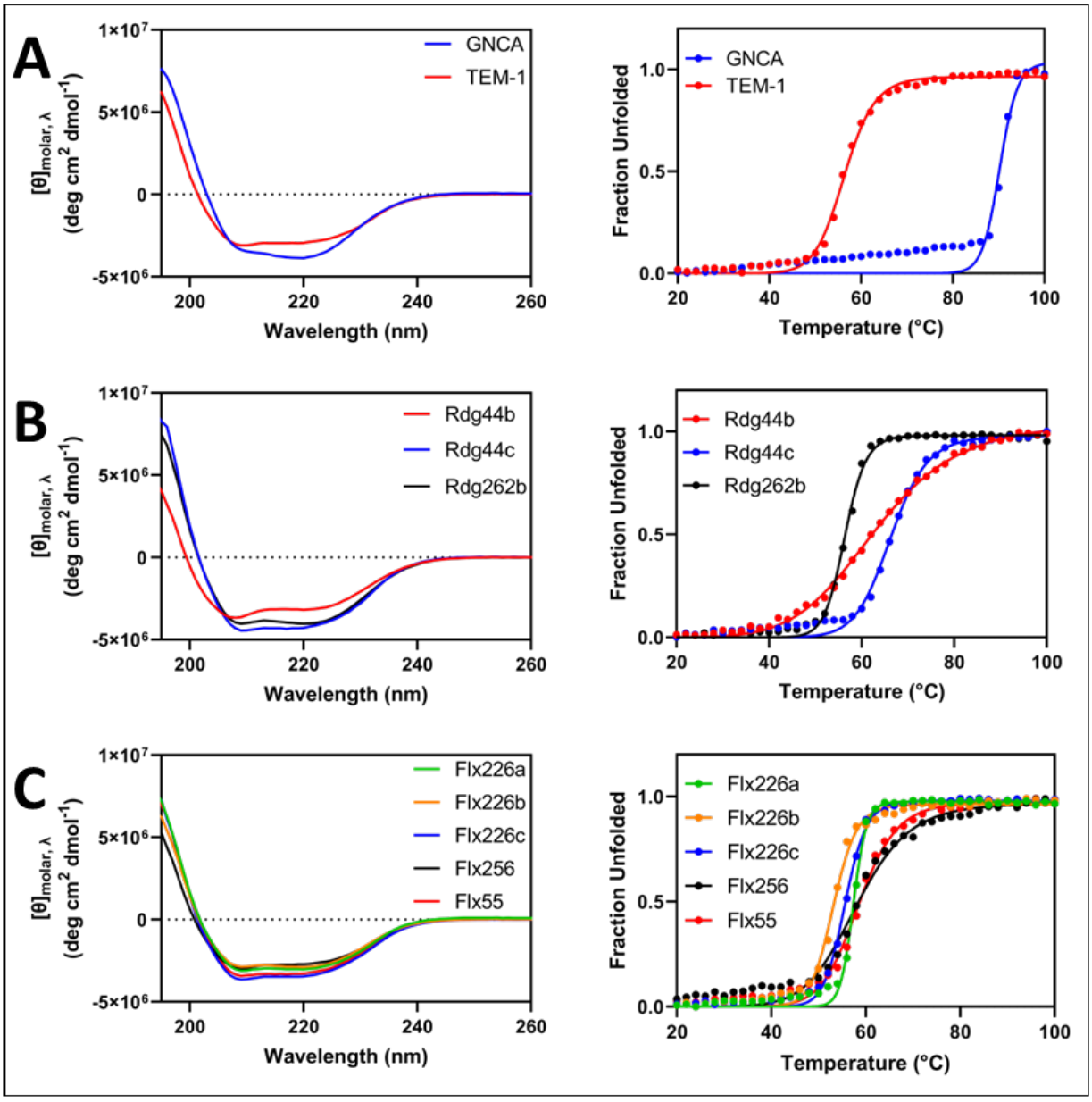


Figure 2.4.4.2 Far-ultraviolet circular dichroism wavelength scans and thermal melts with fitted curves of (A) wild type GNCA and TEM-1 (B) protein designs targeting rigid residues and (C) protein designs targeting flexible residues. All measurements were performed in triplicate on a Jasco J-815 spectrophotometer and adjusted for protein concentration. Thermal melts were monitored by the absorption signal at 222 nm with a temperature slope of 5 °C/min. For wavelength scans and thermal melts, the purified protein was in TBS buffer (10mM Tris 50 mM NaCl, pH 7.0) in a cuvette with a 1 mm path length. Protein concentrations were calculated in triplicate using the absorbance at 280 nm and ranged between 0.18-0.25 mg/mL for all scans.

The residues targeted for design in this study exhibit a broad distribution of distances from the active site. For example, the two rigid residues (V44 and V262) are closer to the active site than any flexible residues that were targeted for design with distances of 10.1 Å and 17.3 Å, respectively, while the distance of the flexible residues from a catalytic residue ranged from 17.5 Å–22.1 Å. We therefore sought to assess whether or not a correlation existed with respect to the distance from a targeted residue to the active site and altered enzymatic function. To do this, we calculated the distances between the Cαs of all residues mutated during the design process and the Cα of the nearest catalytic residue for all experimentally characterized proteins (Table 2) using the PyMOL software (The PyMOL Molecular Graphics System, Version 4.3; Schrödinger, LLC: New York, NY, USA) (*The PyMOL Molecular Graphics System, Version 4.3 Schrödinger, LLC*, n.d.).

The two designed proteins that had the shortest distances between a mutated residue and one of the catalytic residues both targeted residue 262 (Rdg262a and b). Rdg262a carries a mutation at position 233, which is directly adjacent in sequence space to catalytic residue 234. Rdg262b contains the next shortest distance between a mutation and an active site residue at 5.8 Å. Rdg262a showed no activity against ampicillin; it is possible that the observed lack of activity is due to the protein's instability and/or propensity to aggregate as observed during purification. Alternatively, Rdg262b possessed an identical Tm to TEM-1 but showed minimal activity against ampicillin despite containing a mutation that is only ~6 Å away from a catalytic residue. On the other end of the spectrum, the nearest mutations to any catalytic residue in two of the flexible designs, Flx226a and c, are 18.5 and 17.5 Å away, respectively. Both of these

TEM-1 variants showed near native activity against ampicillin, which is consistent with the fact that mutations that are both distant from and uncoupled to the active site should have little effect on activity.

Table 2 Mutations present in the computationally designed proteins and the distance of the nearest mutation to a catalytic residue in angstroms.

Designed Protein	Mutations	Distance from closest catalytic residue (Å)
Rdg44a	E37V, R43P, Y46M, F60Y, R61N, Y264M, T265M, Q278A	11.8
Rdg44b	D35K, E37V, Q39D, R43P, Y46F, F60Y, R61K, I263T, Y264M, T266S, Q278A	11.1
Rdg44c	K32V, D35Q, Q39K, R244I, T265A, R275L, Q278A	9.7
Rdg262a	Y46F, I47Q, E48L, L49T, S59Y, P62A, M182Q, V184K, T188Q, D233C, Y264F	3.9
Rdg262b	Y46F, E48L, P62A, M182T, V184R, I246Q, Y264M	5.8
Flx226a	D254N, I287F	21
Flx226b	A227P, L250F, D254N, K256V, S258T, I287F, W290Y	12.1
Flx226c	A227P, D254N, K256V, I287F	17.5
Flx256	E212A, A227P, A249S, D254N	9
Flx55	E48L, L51A, N52D, S59A, V184R, T188E, T195L, A249M	9.8

In the remaining designs, the distribution of distances between the nearest catalytic residue and a designed mutation are much more similar irrespective of whether rigid or flexible residues were targeted. For example, Rdg44a and Flx226b both have mutations

that are 12.1 Å from a catalytic residue and Rdg44c and Flx55 have mutations that are 9.7 Å and 9.8 Å away from the catalytic residues, respectively. As these pairs of proteins contain one rigid and one flexible design and also exhibit similar distances between the nearest mutation and any catalytic residue, they appear to provide a direct test of the implications of targeting mutations to flexible vs. rigid residues. Interestingly, Rdg44a was highly unstable and aggregation prone despite only having mutations over 10 Å away from the catalytic residues. In contrast, Rdg44c had activity against ampicillin that was three orders of magnitude less than the wild type protein, but also showed a 10 °C increase in T_m relative to TEM-1. Alternatively, both flexible designs (Flx226b and Flx55) maintained substantial activity against ampicillin and exhibited T_ms that were within 3 °C of wild type TEM-1 (Table 1). These data further support the notion that rigid, highly coupled residues play a large role in determining both the activity and physical properties of TEM-1. Furthermore, the fact that the rigid designs that adopted a native-like fold showed a substantial decrease in activity supports the notion that our dci metric can provide meaningful information regarding residues that may be able to affect protein function via allosteric dynamic coupling to the active site.

2.4.5 Dynamics Analysis of the Designed Proteins. Experimental characterization of our designed proteins demonstrated that the MIC_{amp} values of the rigid designs were significantly reduced relative to both TEM-1 and the flexible designs irrespective of the distances between the nearest mutations and the catalytic residues. This suggests that changes in the local network of interactions surrounding rigid residues that exhibit long-range dynamic coupling with the active site may allosterically alter the flexibility of active site residues. To further analyze this possibility using our computational metrics,

we calculated the flexibility of the active site residues in both sets of designed proteins using the dfi metric. The dfi values of each catalytic residue in our experimentally characterized proteins were subtracted from those of TEM-1 to generate a Δdfi profile (Figure 2.4.5.1.A).

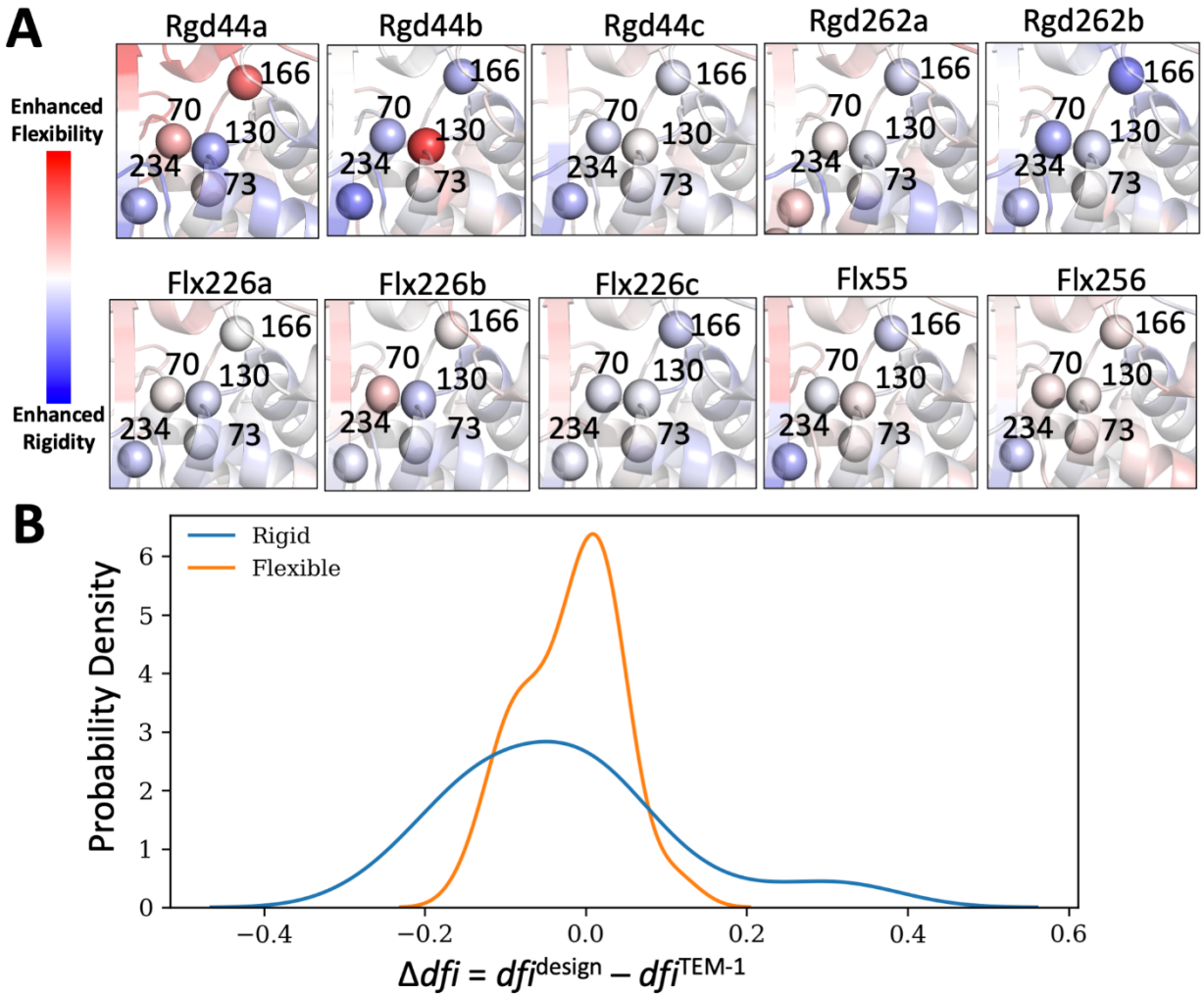


Figure 2.4.5.1. The change in dynamics as measured by the Δdfi mapped onto the catalytic residues of each experimentally characterized protein. A) Catalytic residues are modeled as spheres and color coded by their change in dfi score relative to TEM-1. B) The Δdfi distribution of active site residues in the flexible and rigid designs. The flexible design distribution shows a low variance compared to that of the rigid designs. A change in dfi score of + 0.2 is noteworthy as it is indicative of a shift in flexibility. This analysis suggests that designing new interactions around a rigid residue that is dynamically coupled to the active site can allosterically modulate the flexibility/rigidity of the amino acids in the active site.

A clear difference between the Δdfi values of the catalytic residues of the rigid and flexible designs was observed (Figure 2.4.5.1). Namely, the catalytic residues in the rigid designs underwent a greater change in relative flexibility (both increases and decreases) compared to the flexible designs. Alternatively, the relative flexibilities of the catalytic residues in the flexible designs exhibited a narrower distribution centered at zero (Figure 2.4.5.2). These data support the notion that the rigid residues we chose are highly coupled to the active site (as suggested by our original dci analysis) and also that targeting the local interaction of allosteric rigid residues can indeed alter the flexibilities of residues, even if they are separated by substantial distances.

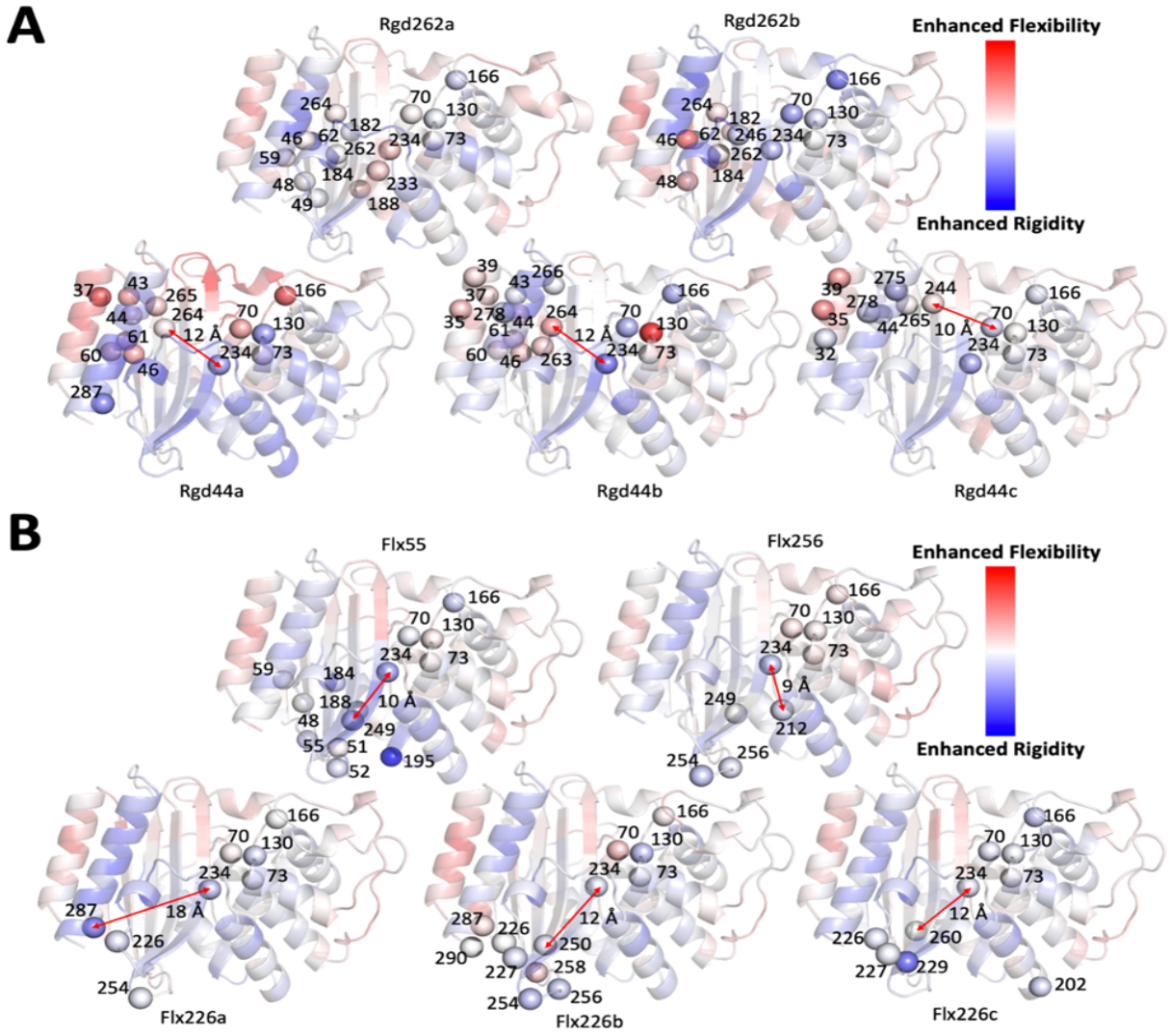


Figure 2.4.5.2. The change in the dynamics profiles of experimentally characterized rigid (A) and flexible (B) designs (Δdfi values) are mapped onto the TEM-1 structure. Point mutations around the residues targeted for design and the catalytic residues in TEM-1 are shown as spheres and labeled with their residue indices. The distance between the mutations closest to the catalytic residues are marked with red arrows and labeled with the corresponding distance in angstroms. The minimum distance in most designs is larger than 10 Å (Rgd262a and b and Flx256 are exceptions), which suggests that the changes in dynamics of catalytic residues is due to distal allosteric communication with the active site in many instances.

Our experimental results and the detailed dfi profiling of the experimentally characterized designs brought to light the fact that our initial PCA analysis did not appear to adequately discriminate between the activities of the designed proteins. Although

designs in which rigid, coupled residues were targeted often possessed vastly different properties than those in which flexible, uncoupled residues were targeted, many of these designs clustered in similar areas of the PCA (Figure 2.4.3.2). Therefore, we sought to develop a new metric that might have a greater discriminatory ability than the PCA alone. We therefore developed an iterative method that we have termed the Dynamic Distance Analysis (dda) in which the "dynamic distance" of a designed protein to either TEM-1 or GNCA is computed relative to those of randomly selected groups of designed proteins. As the distance between any two proteins in a PCA (based on their three principal eigenvectors, see Methods and Materials) depends on the component proteins used to generate that PCA, randomly selected sets of designed proteins should yield a much better picture of the true relationship between a given designed protein and a target protein (TEM-1 and GNCA).

To generate the dda profiles of our designed proteins, we used a bootstrapping approach in which we first generated multiple PCAs using small, randomly chosen subsets of designed proteins and then iteratively measured the distances between the dfi profiles of each designed protein and both GNCA and TEM-1 (Figure 2.4.5.3). When we clustered the dda profiles of the rigid and flexible designs using a new PCA; a clear separation between the two emerges (Figure 2.4.5.3), which correlates well with their biophysical characterization. For example, flexible designs Flx55 and Flx256 cluster together in our dda analysis and also possess similar MIC_{amp} values (750 µg/mL). Similarly, Flx226a and Flx226c, whose MIC_{amp} values are the same as TEM-1 (1500 µg/mL), also appear in very similar regions of the dda PCA. The two rigid designs, Rgd44a and Rgd262a, which exhibited aggregation during purification, are both found as

outliers in the dda clustering. Notably, Rgd44c and Rgd262b, which exhibit higher thermostabilities and similar MIC_{amp} values to TEM-1, are also clustered in the same vicinity.

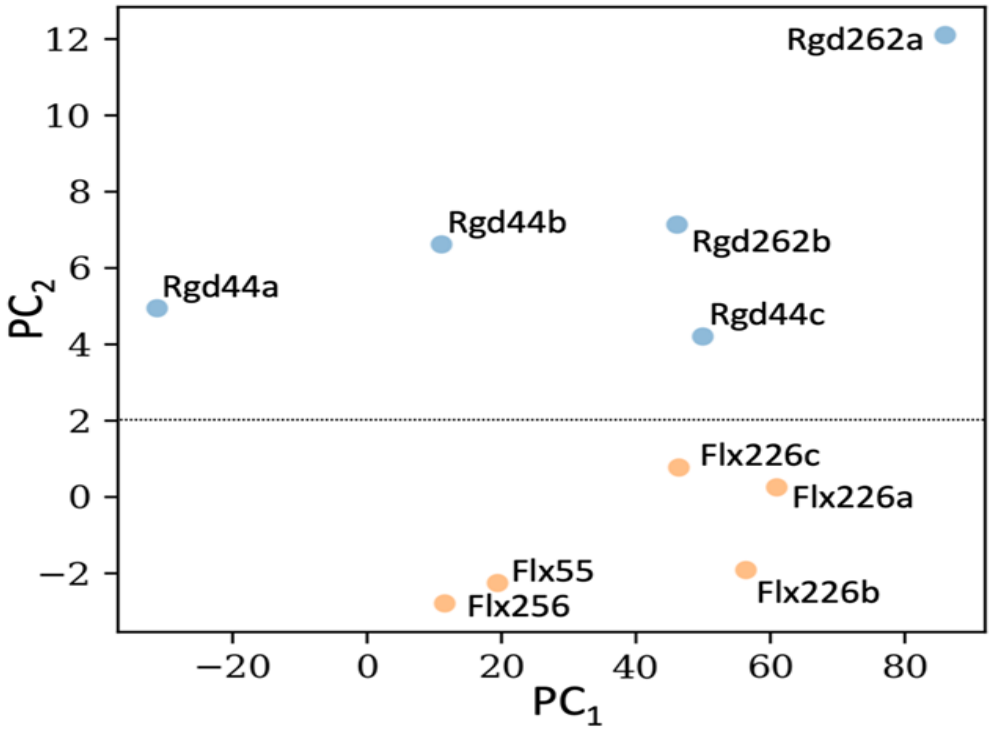


Figure 2.4.5.3 The dynamic distances are clustered for all characterized allosteric rigid (blue) and uncoupled flexible (orange) designs. The weights of PC1 and PC2 are 250 and 30, respectively. The rigid designs and the flexible designs cluster separately. Designed proteins with similar MIC_{amp} values, (Flx55 and Flx256), (Flx226c and Flx226a), (Rdg262b and Rdg44c) cluster in the same vicinity.

In an effort to assess whether or not the trends observed in the dda analyses of experimentally characterized proteins were universal, we applied dda to all the designed proteins, even those not chosen for characterization. Interestingly, the dynamic distances of the rigid designs are biased away from TEM-1 relative to their flexible design counterparts (Figure 2.4.5.4); conversely, the flexible designs form a narrower distribution that is closer to TEM-1. This suggests that flexible residues that are not

coupled to the active site do not likely contribute to the collective motion of the protein as substantially as do rigid residues. When the distances of our designed proteins to GNCA are considered, the uncoupled flexible designs display a sharp, narrow distribution that is distant from GNCA (Figure 2.4.5.4). Alternatively, the distribution of the rigid designs is broad and contains proteins with dynamic profiles that are more like that of GNCA. These data suggest that the re-design of the environment surrounding rigid residues appears to alter the dynamics of TEM-1 more substantially than when the environment surrounding uncoupled flexible residues is targeted.

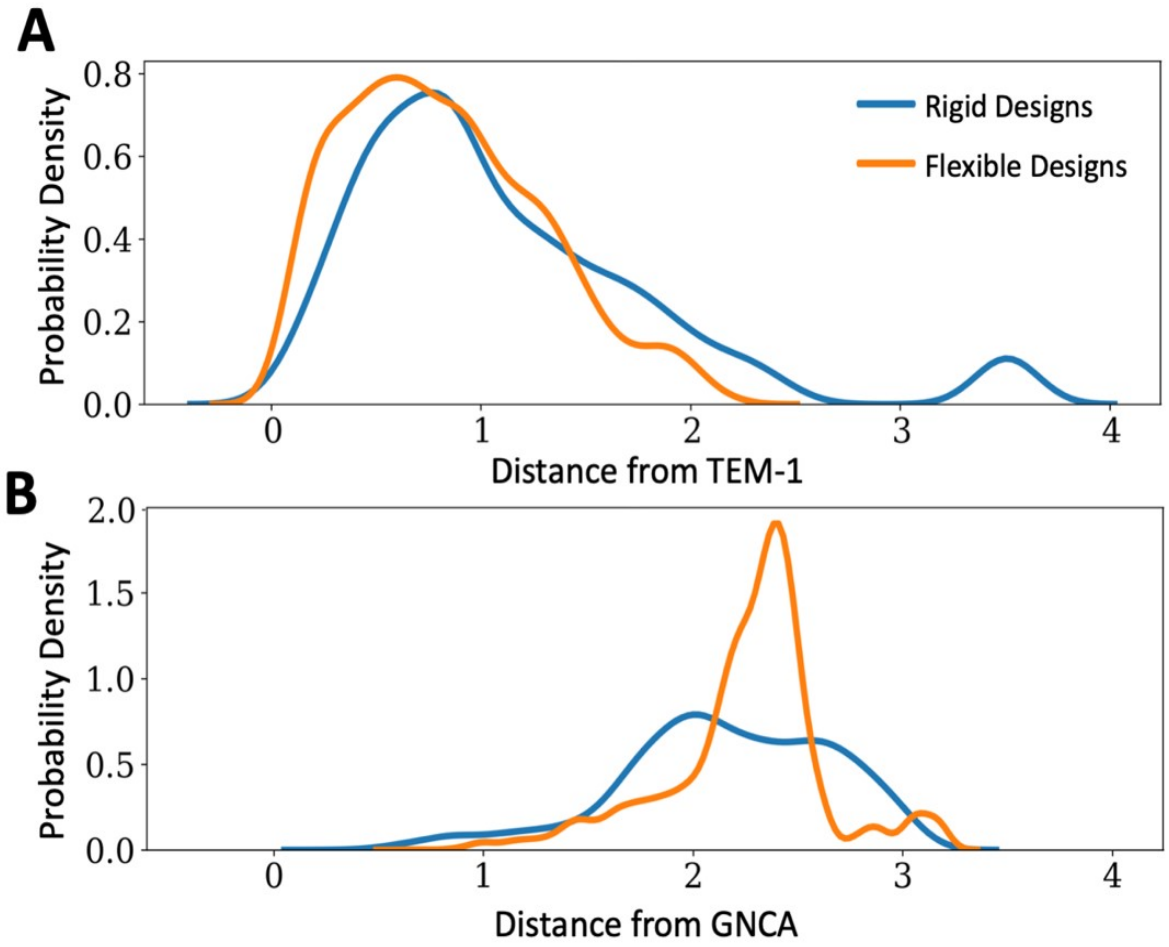


Figure 2.4.5.4. Dynamic distance distribution from (A) TEM-1 and (B) GNCA for all experimentally characterized rigid (blue) and flexible designed proteins (orange). The distribution of the rigid designs shows a displacement moving away from TEM-1 and

closer to GNCA. Inversely, the uncoupled flexible designs form a narrow distribution close to TEM-1 and further away from GNCA.

2.5 Conclusions

The goal of this work was to better understand the relationship between structure and function in the TEM family of β-lactamases. Building on previous evolutionary studies on the β-lactamase enzyme TEM-1 (Zou et al., 2015), we explored the hypothesis that rigid residues can serve to both establish the global dynamic profile of the enzyme and exert substantial influence over physical properties (e.g., substrate specificities) so long as long-range coupling exists between the rigid residues and the active site. To explore this, we used the Rosetta computational protein design software to re-design the local network of interactions surrounding residues that fit the aforementioned criteria. Our designed proteins were analyzed using computational metrics that assessed both the global dynamic profile and the allosteric coupling of each residue to the active site. Based on these metrics, a subset of our designed proteins was selected for experimental characterization.

Ten designed TEM-1 variants were characterized with respect to the minimal inhibitory concentration of ampicillin as well as thermostability. These data suggested that targeting mutations to environments surrounding rigid residues that were highly coupled to the active site often resulted in a substantial shift in protein stability and function; alternatively, targeting flexible, uncoupled residues resulted in protein variants with more native-like activities and thermostabilities. Namely, when mutations were targeted to the vicinity of two rigid residues that do not directly interact with the active site, but which are highly coupled to it, a substantial reduction in TEM-1′s ability to

degrade its native substrate was observed in all cases even though native-like folds were maintained in many cases. Alternatively, thermostabilities and activities against TEM-1′s native substrate were maintained in a set of designed proteins in which residues that were neither rigid nor predicted to be coupled to the active site were targeted for mutagenesis. These results are consistent with our computational analyses of the designed proteins' dynamics. Namely, it appears that altering the local interactions surrounding rigid residues that are highly coupled to the active site can allosterically alter the flexibility profiles of active site residues at a distance, which can in turn alter the biophysical properties of the enzyme. In an effort to identify an analytical method that was more informative as to the activities that designed proteins might possess, we developed a novel metric that measures the "dynamic distance" between two proteins. Many of our designed proteins with similar functional properties were observed to cluster together when analyzed by this algorithm. These results not only further support the potential importance of mutations in the vicinity of rigid residues, but also support the fact that coupling between distal residues and the active site can have profound effects on enzyme activities.

The relationship between protein dynamics and function is highly complex and studying it represents an exceedingly difficult challenge (Ma et al., 2011; Maier et al., 2015; Orencia et al., 2001; Salverda et al., 2010; M. K. Singh & Dominy, 2012; Zhang et al., 2020). Our approach represents a new method for exploring this subject in a highly directed manner. We hope that additional application of these methods to distinct residues in TEM-1 will ultimately provide a more complete understanding of the complex dynamic landscape present in this class of proteins. This could not only facilitate a rapid

prediction of the biochemical properties of new clinical isolates but could also pave the way for the development of new antibiotics that specifically target new protein conformations accessible only through alterations of the global dynamic profile. Finally, the methods reported here could also find use in understanding the dynamic profiles of other enzyme classes, which could have profound implications from the perspective of understanding and treating diseases.

2.6 Acknowledgements

This research was funded by The National Science Foundation, grant number 1901709. The authors thank Jose Sanchez-Ruiz (Universidad de Granada) for the generous gift of the GNCA expression plasmid and Ron Mills for helpful review of the manuscript.

CHAPTER 3

STRUCTURAL INSIGHTS INTO TEM-1 VARIANTS

3.1 Abstract

To identify if the functional and stability changes observed in the TEM-1 variants from the previous study were due to structural changes introduced by the mutations, Flx226a, Flx226b, Flx226c, Rdg44c and Rdg262b were chosen for crystallization. These variants were chosen because they represent a wide range of function against ampicillin as well as a wide range of melting temperatures as measured by CD spectroscopy. Models of the variants, solved using molecular replacement, were compared to wild type TEM-1 to see if any of the variants had conformational changes in the active site architecture. Aside from the presence of a pH dependent acetate adduct on Ser70 in all the models, the main catalytic residue, the active site architecture was unremarkable. One residue, Tyr105 that delineates one of the sides of the active site was observed in two conformations, an "inward" and an "outward" conformation that was independent of catalytic efficiency. Overall, the models did not provide obvious reasons for differences in function amongst the variants. However, the observation of the pH dependent serine acetate adduct, has to my knowledge, never been reported and is potentially of interest to industrial applications where serine hydrolases are used in low pH conditions.

3.2 Introduction

In Chapter 2, we explored how changing networks of interactions around residues calculated to be rigid and highly coupled to the active site would affect enzyme function in TEM-1. To this end, we employed RosettaDesign to make mutations around target residues and characterized chosen designs for changes in function using MIC_{amp} assays

and for changes in stability using thermal melts. When we designed around rigid residues, enzyme function and thermostability was substantially affected. However, when we designed around the flexible residues, native function was basically maintained and melting temperatures were similar to wild type.

To identify if the mutations made to the variants caused structural changes, Flx226a, Flx226b, Flx226c, Rdg44c and Rdg262b were chosen for crystallization attempts. These variants were chosen because they represent a wide range of function against ampicillin as well as a wide range of melting temperatures (Table 1). Flx226a and Flx226c maintained wild type catalytic efficiency and thermostability as measured by MIC_{amp} assays (1500 μg/mL) and CD spectroscopy. However, the third flexible variant, Flx226b, was an order of magnitude less efficient against ampicillin (375 μg/mL) when compared to wild type TEM-1. In addition, the melting temperature of Flx226c was 3 ºC lower than wild type which could have been a factor in the loss of function. Rdg44c and Rdg262b, the two designs where rigid residues were targeted, had a significant loss of function, maintaining only ~2% of wild type activity (26 μg/mL). However, the thermostability of Rdg262b was unaffected by the mutations while the melting temperature of Rdg44c went up by 10 ºC. It was these differences in both function and thermostability that lead me to believe that there might be structural changes in the variants.

3.3 Materials and Methods

3.3.1 X-ray Crystallography. Variants were expressed and purified following the procedure outlined in Chapter 2 Materials and Methods. After verifying purity using a 12% SDS PAGE gel, a solution of Flx226a (10mM Tris 50 mM NaCl, pH 7.0) was

divided in half to determine the crystallization conditions that would lead to diffraction quality crystals. One sample was concentrated to A_{280} ~ 9 mg/mL and the other was concentrated to A_{280} ~ 20 mg/mL for high-throughput crystal condition screenings. Protein concentrations were obtained using a NanoDrop visible spectrophotometer (Thermo Scientific, Waltham, Massachusetts). A Mosquito crystallization robot (SPT Labtech, Melbourn (Cambridge), UK) performed primary screening of crystallization conditions against three hanging drop vapor diffusion Hampton Research crystal screening libraries (Index HT, SaltRx HT, and PEGRx HT). Each screen contained 96 conditions and each condition was tested three times with different v/v ratios of protein to reservoir drop for a total of 864 conditions for each protein concentration in 100, 200, and 300 nL drop sizes. The plates of conditions around the proteins at 9 mg/mL were incubated at room temperature while the plates of conditions around the proteins at 20 mg/mL were incubated at 4 °C. The condition that produced crystals after 7 days was screened using larger-volume hanging drop vapor diffusion to identify the optimal pH, buffer-to-protein ratio, and concentration of the cryoprotectant, polyethylene glycol (PEG) 300, to grow large, single crystals. Diffraction-quality crystals were grown in similar conditions at room temperature for the five variants at a concentration ~ 9 mg/mL. Flx226a, Flx226c and Rdg44c crystals were grown in 0.1 M sodium acetate pH 4.9, 45% PEG 300. Flx226b and Rdg262b crystals were grown in 0.1 M sodium acetate pH 4.9, 50% PEG 300. Each drop contained 2 µL of protein solution mixed with 2 µL of reservoir solution. Crystals in the shape of tetragonal bipyramids were grown until no new growth was apparent which was ~ 1.5 weeks (Figure 3.2.1).

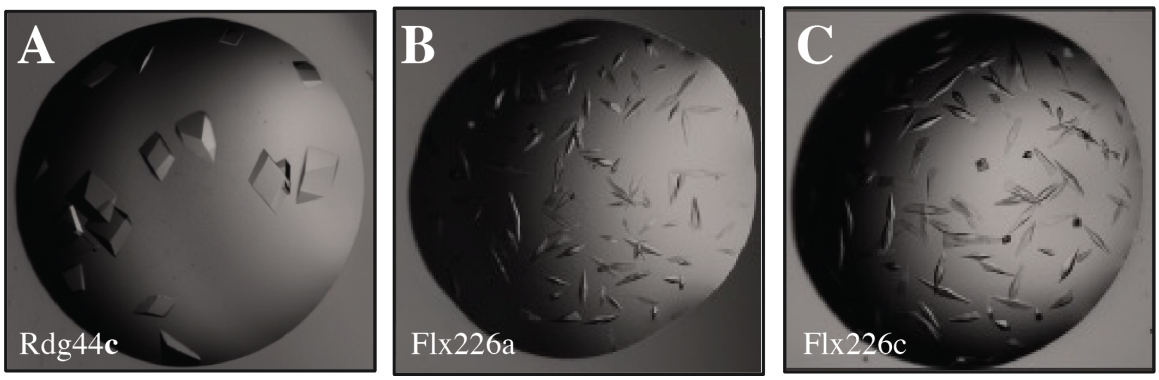


Figure 3.3.1 Photographs of three hanging drops of protein crystals. (A) Crystals of Rdg44c. (B) Crystals of Flx226a. (C) Crystals of Flx226c. All of the TEM-1 variants crystallized as tetragonal bipyramids.

Initially, crystals for Rdg262b grew in a starburst geometry from a single point which was suboptimal for diffraction. Therefore, one "starburst" crystal was crushed and used to seed fresh solutions of a 1:1 ratio of protein to mother liquor using a streak seeding technique with a cat whisker. New diffraction quality crystals for Rdg262b grew in ~2 days.

To obtain a dataset of Flx226b at physiological pH, left-over crystals from successful initial crystallization attempts were harvested and soaked for 30-60 minutes in increasing (v/v) concentrations of 100 mM Tris-HCl pH 7.5, 50% PEG.

All crystals were harvested, and flash frozen in liquid nitrogen prior to data collection at 100 K. Data was collected at the Stanford Synchrotron Radiation Lightsource (SSRL) operated for the U.S. Department of Energy by Stanford University. Data were indexed, refined, integrated, and scaled using the XDS software package (Kabsch, 2010a, 2010b). All structures were solved by molecular replacement with the Phaser 2.7.17 software using PDB ID: 1btl as the search model (McCoy et al., 2007). Model building took place using the program Coot (Emsley et al., 2010). All models

were refined using Refmac5 which is part of the CCP4 software package (Kovalevskiy et al., 2018; Murshudov et al., 1997, 2011; Nicholls et al., 2018; Winn et al., 2011).

3.3.2 Mass Spectrometry of Flx226a and Rdg44c. Protein solutions left-over from previous crystallization attempts of Flx226a and Rdg44c were centrifuged at 13,000x *g* for 10 minutes at 4 °C to remove any precipitated protein from the solution. Solutions were then divided in half. One sample was dialyzed overnight at 4 °C into 10 mM sodium acetate pH 4.5, 50 mM NaCl. The second sample was dialyzed overnight at 4 °C into 10 mM Tris pH 7.0, 50 mM NaCl. Samples were diluted to ~50 μg/mL in 18 MΩ water. Data were collected via direct injection on an Agilent 6530 Accurate-Mass Q-TOF LC/MS instrument. Deconvolution of the spectra was done with the Agilent Masshunter Bioconfirm software package.

3.4 Results and Discussion

Crystal structures for the TEM-1 variants were solved to the following resolutions. Flx226a and Flx226c were solved to 1.52 Å and 1.75 Å, respectively. Flx226b (pH 4.9) and Flx226b (pH 7.5) were solved to 1.25 Å and 1.75 Å, respectively. Rdg44c and Rdg262b were solved to 1.04 Å and 1.53 Å, respectively (Table 3).

3.4.1 Electron Density Around Ser70 The active site architecture was of immediate interest when comparing the models built from crystal structure data of the variants. I hypothesized that when we designed around the highly coupled rigid residues that changes in structure may have been propagated to the active site. These changes may have been as subtle as a change in catalytic residue side chain conformation or as grand

Table 3 Crystallization statistics for the TEM-1 variants

Structure	Rdg44c	Rdg262b	Flx226a	Flx226b (pH 4.9)	Flx226b (pH 7.5)	Flx226c
Experimental	0.1 M sodium acetate pH 4.9 45% PEG 300	0.1 M sodium acetate pH 4.9 50% PEG 300	0.1 M sodium acetate pH 4.9 45% PEG 300	0.1 M sodium acetate pH 4.9 50% PEG 300	0.1 M Tris-HCl pH 7.5 50% PEG 300	0.1 M sodium acetate pH 4.9 45% PEG 300
Data collection						
Space group	P 32 2 1	P 41 21 2	P 4 2 2	P 4 2 2	P 4 2 2	P 4 2 2
Cell dimensions a, b, c (Å) α, ß, γ (deg)	63.612 63.612 126.928 90 90 120	46.552 46.552 257.558 90 90 90	46.565 46.565 257.7 90 90 90	46.67, 46.67, 257.434, 90, 90, 90	47.078, 47.078, 257.168, 90, 90, 90	46.508 46.508 257.099 90 90 90
Total Reflections	1624602	581225	637440	1071646	182556	971156
Unique Reflections	128981	43723	44236	70855	30291	57640
Resolution (Å)	33.5552 - 1.0407	50.000 - 1.529	46.565 - 1.52463	46.67 - 1.25	47.078 1.74807	46.508 - 1.32735
I/σ(I)	26.1	12.1	9.2	31.3	13.5	10.3
R_{meas}	0.028	0.097	0.123	0.026	0.033	0.12
R_{pim}	0.008	0.026	0.033	0.007	0.014	0.029
CC1/2	0.836	1.321	1.774	0.607	1.348	1.812
Completeness (%)	90.29	98.1	98.28	87.84	98.46	85.87
Redundancy	12.6	13.3	14.4	15.1	6	16.8

as an alternative backbone placement. Notwithstanding, our observations of the electron density maps of the variant active sites were originally befuddling. In the electron density maps of every variant, we observed a strong electron density peak centered around the main catalytic residue, Ser70. This electron density was tetrahedral in shape extended from the side chain of Ser70 to a lobe ~1.5 Å from Ser70 Oγ. Because the wild type TEM-1 model has conserved two waters in the active site 2.7 and 2.9 Å away from Ser70 Oγ, I reasoned that the density was due to the presence of extra waters. Therefore, first attempts at building the active site model focused on the addition of 2-3 waters ~2.6 Å from Ser70 Oγ. After refinement with Refmac5, there was still a strong positive difference peak extending from Ser70 Oγ to a sphere 1.5 Å away and centered between the Ser70 Oγ and two modeled water molecules (Figure 3.4.1.1.A). It was evident that the electron density near Ser70 was not simply extra waters. The distance between the density peak and Ser70 Oγ prompted us to explore the possibility that a small molecule either in the cytosol, purification buffers, or in the mother liquor must have reacted with

Ser70 to form a tetrahedral adduct. The simplest explanation was that the small molecule was present in the crystallization conditions.

The crystallization buffer that uniformly gave rise to diffraction quality crystals was 100 mM sodium acetate (pH 4.9) with PEG 300, but it seemed unlikely that Ser70 would react with an acetate ion. However, at low pH, an activated Ser could make a nucleophilic attack on a neutral acetic acid molecule. We compared the differences in R_{values} post refinement when we solely modeled waters and when we modeled waters with an acetate adduct in the active site. When three waters were modeled the R_{values} were 0.1254 and 0.1405 for R_{factor} and R_{free}, respectively. When Ser70 was modeled as an acetate adduct at 60% occupancy and the three waters at 40% occupancy, the R_{values} were 0.1228, 0.1384 for R_{factor} and R_{free}, respectively (Figure 3.4.1.1.B).

This observation gave us confidence in the presence of an acetate adduct. However, we decided to investigate further with wet lab experiments.

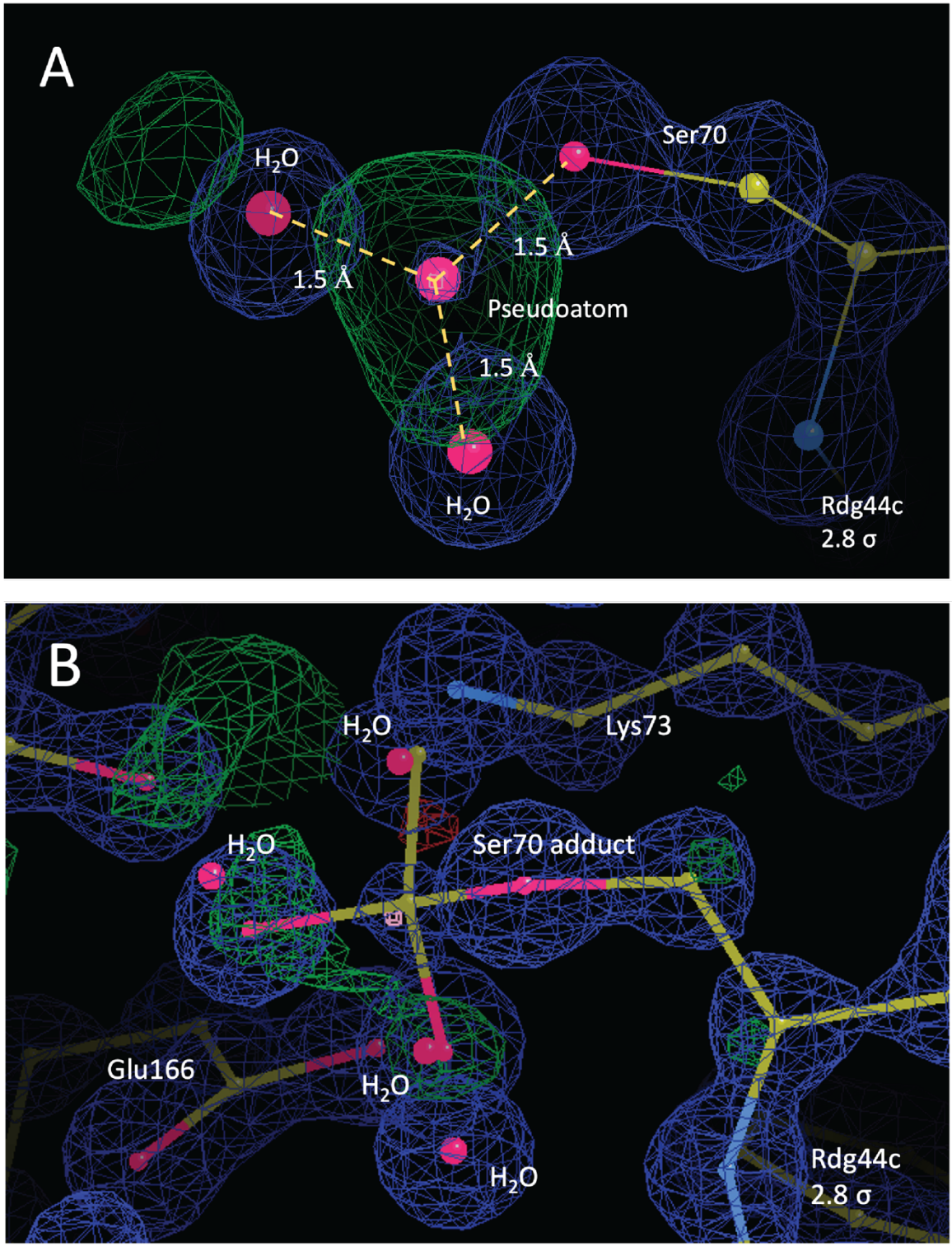


Figure 3.4.1.1 Electron density and stick models of the Rdg44c active site at 2.8 σ. (A) When three water molecules are modeled in the strong electron density surrounding Ser70, a strong difference peak appears 1.5 Å away from Ser70 Oγ indicating that there is

a covalent bond. In the image, a pseudoatom was placed in the center of the green density for distance measurements. (B) The difference peak is not present when a Ser70 adduct is modeled at 60% occupancy. Three water molecules are modeled at 40% occupancy for the lowest R_{values} after refinement with Refmac5. Other catalytic residues are modeled as sticks.

To further explore the presence of a Ser70 adduct, we took older purified protein samples of Flx226a and Rgd44c that were left over from crystallization studies and prepared them for mass spectrometry analysis. We divided both samples into two aliquots. The first sample was dialyzed overnight at 4 °C into 10 mM sodium acetate pH 4.5, 50 mM NaCl. The second sample was dialyzed overnight at 4 °C into 10 mM Tris pH 7.0, 50 mM NaCl. Both samples were then diluted in 18 MΩ water for input into the mass spectrometer. We reasoned that we could obtain the difference in mass to back calculate the molecular weight of the addition to Ser70. Unfortunately, data from the mass spectrometer did not show any evidence of an adduct and the masses were spot-on for both proteins regardless of dialysis buffer.

Because the crystallization conditions were at low (pH 4.9) we reasoned that the adduct could be pH dependent which is why it was not observed in mass spectrometry experimental data taken at physiological pH. Therefore, our next step to solving the adduct mystery was to collect data on one of the leftover crystals in a buffer at higher pH. We progressively soaked a selection of Flx226b crystals in increasingly higher pH buffers until it was stable in 100 mM Tris-HCl pH 7.5, 50% PEG. This buffer exchange not only raised the pH of the buffer, but also removed the acetate ions from the solution. We collected diffraction data on this crystal from the SSRL and solved the crystal structure for Flx226b at pH 7.5. To our delight, the electron density in the active site of

this variant at pH 7.5 was vastly different than that observed at pH 4.9. The electron density clearly showed evidence of the two conserved waters 3.2 and 2.8 Å from Ser70 Oγ. There was no evidence of an adduct around Ser70, lending weight to our hypothesis that the Ser70 acetate adduct was pH dependent.

3.4.2 Analysis of the Variant Active Sites Once the mystery of the adduct was solved, analysis focused on potential changes in the active site architecture that might lead to changes in function against ampicillin. Superimposing the models of the crystallized variants with PDB ID: 1btl showed no obvious differences in backbone conformation or side chain conformations between the variants with serious losses of function to variants that maintained function. All catalytic residues were in the same basic conformation as those observed in the 1btl model (Figure 3.4.2.1).

The only observed differences in catalytic residue side chain conformation was in the models of Flx226b. At pH 7.5, the torsion angle around Ser130 Cβ was observed to be rotated 66 degrees toward Lys234. No other side chains in the active site were affected by this slight rotation (Figure 3.4.2.2). At pH 4.9, with the Ser70 adduct present, the torsion angle around Ser130 Cβ adopts the same basic conformation as the wildtype (1btl) model (within ~15 degrees). However, the 4.9. the torsion angle around Lys234 Cε was observed to be rotated ~75 degrees toward the backbone of I127 to form a sidechain-backbone hydrogen bond (Figure 3.4.2.2). There's no reason to believe that the Lys234 rotation in Flx226b is due to the presence of the adduct since the other variants with a Ser70 adduct at pH 4.9 are observed with Lys234 in same conformation as wild type TEM-1. In all the models aside from Flx226b, Lys234 Nζ was observed within 2.8 Å

(hydrogen bonding distance) of Ser130 Oγ, the backbone carbonyl of Ser235, and a conserved water.

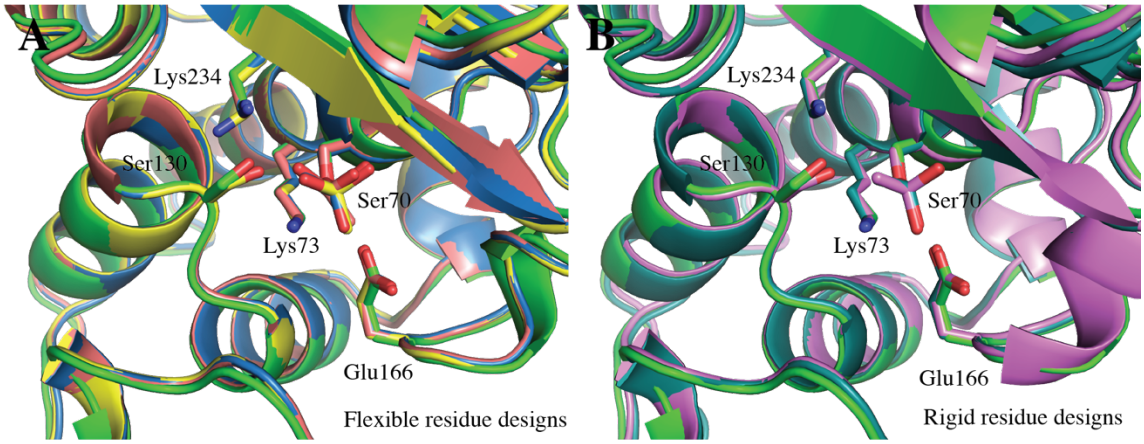


Figure 3.4.2.1 Models of the active site of the crystallized TEM-1 variants at pH 4.9. (A) Models of the flexible designs, Flx226a (salmon), Flx226b (yellow), and Flx226c (blue) superimposed with wild type TEM-1 (PDB ID: 1btl) (green). The main catalytic residues are modeled as sticks. Aside from the torsion angle around Lys234 Cε rotation of ~75 degrees toward the backbone of I127, (not shown) to form a sidechain-backbone hydrogen bond, no significant changes in side chain conformation with respect to wild type TEM-1 are observed in the models. (B) Models of the rigid designs, Rdg44c (teal) and Rdg262a (violet) superimposed with wild type TEM-1 (PDB ID: 1btl) (green). The main catalytic residues modeled as sticks. No significant changes in side chain conformation with respect to wild type TEM-1 are observed in the models.

3.4.3 Analysis of Tyr105 Tyr105 is a Class A conserved residue that delineates one of the edges of the active site wall and has been identified as important to substrate recognition and stabilization (Doucet et al., 2004, 2007b).This stabilization is believed to be due to a stacking interaction between the aromatic ring of Tyr105 and the thiazolidine ring on penicillin-derived antibiotics (Doucet et al., 2004). The proximity of the aromatic ring of Tyr105 to the thiazolidine ring of penicillin-based substrates is apparent in an x-ray crystal structure model at 1.7 Å resolution of TEM-1 in complex with benzyl penicillin (PDB ID: 1fqg) with an E166N mutation preventing it from deacylating the

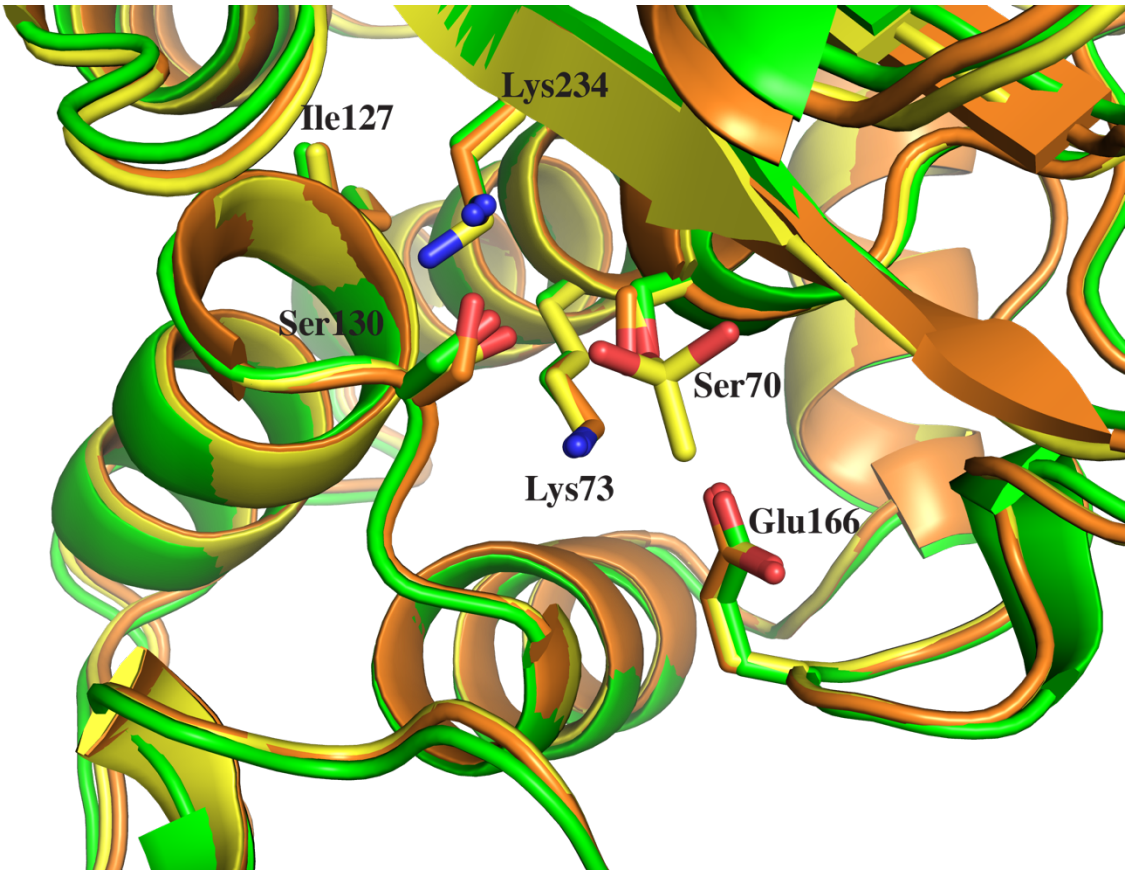


Figure 3.4.2.2 Models of the Flx226b active site at pH 4.9 (yellow) and pH 7.5 (orange) superimposed with wild type TEM-1 (PDB ID: 1btl) (green). At pH 4.9, Ser70 is modeled with an adduct and the torsion angle around Lys234 Cε is observed rotated ~75 degrees toward the backbone of I127 to form a sidechain-backbone hydrogen bond. At pH 7.5, the torsion angle around Ser130 Cβ is rotated 66 degrees toward Lys234.

acyl-enzyme intermediate in the reaction (Sielecki et al., 2003). In this model, Tyr105 has a torsion angle around the Cβ atom of 63.7 degrees and the phenol is pointing toward Met129. In this "inward" conformation, Tyr105 Oγ is within hydrogen bonding distance (3.0 Å) of the backbone carbonyl of Met129. Additionally, the Tyr105 aromatic ring is in an edge-to-face stacking interaction with Pro107 as well as within the ideal sulfur-π interaction distance of 4.4 Å of the sulfur atom on the penicillin thiazolidine ring (Ringer et al., 2007). Notably, the conformation of Tyr105 in the wild type model (PDB ID: 1btl)

is the same as that observed in the model of TEM-1 with an acyl-intermediate (PDB ID: 1fqg).

Previous studies have found that in a co-crystal structure of a TEM-1 homologue derived from *Staphylococcus aureus* named PC1, bound to an inhibitor, *p*-nitrophenyl[[*N*-(benzyloxycarbonyl)amino]methyl]phosphonate, (PDB ID: 1blh), Tyr105 will adopt an "outward" conformation that is not observed in the wild type TEM-1 structure. Nor is this "outward" conformation observed in the apo TEM-1 crystal structure as previously discussed. With the bound inhibitor, the Tyr105 torsion angle around Cβ is 179.4 degrees so that its phenol points toward solvent (C. C. H. Chen et al., 1993). It has been hypothesized that the Tyr105 outward conformation is due to the presence of the inhibitor. It should be noted that the structure of the methylphosphonate inhibitor covalently bound to Ser70 does not preclude the "inward" Tyr105 conformation observed in the apo-TEM-1 structure and in the TEM-1 E166N structure co-crystallized with penicillin G. The methylphosphonate inhibitor also has a smaller chemical footprint than the acylated penicillin G structure. Additionally, when models of PC1 (PDB ID 1b1h), and TEM-1 E166N bound with penicillin G are superimposed, the distance between the center of the Tyr105 aromatic ring in the outward-facing conformation to the sulfur atom on the thiazolidine ring on penicillin is 4.7 Å, also well within sulfur-π interaction distance.

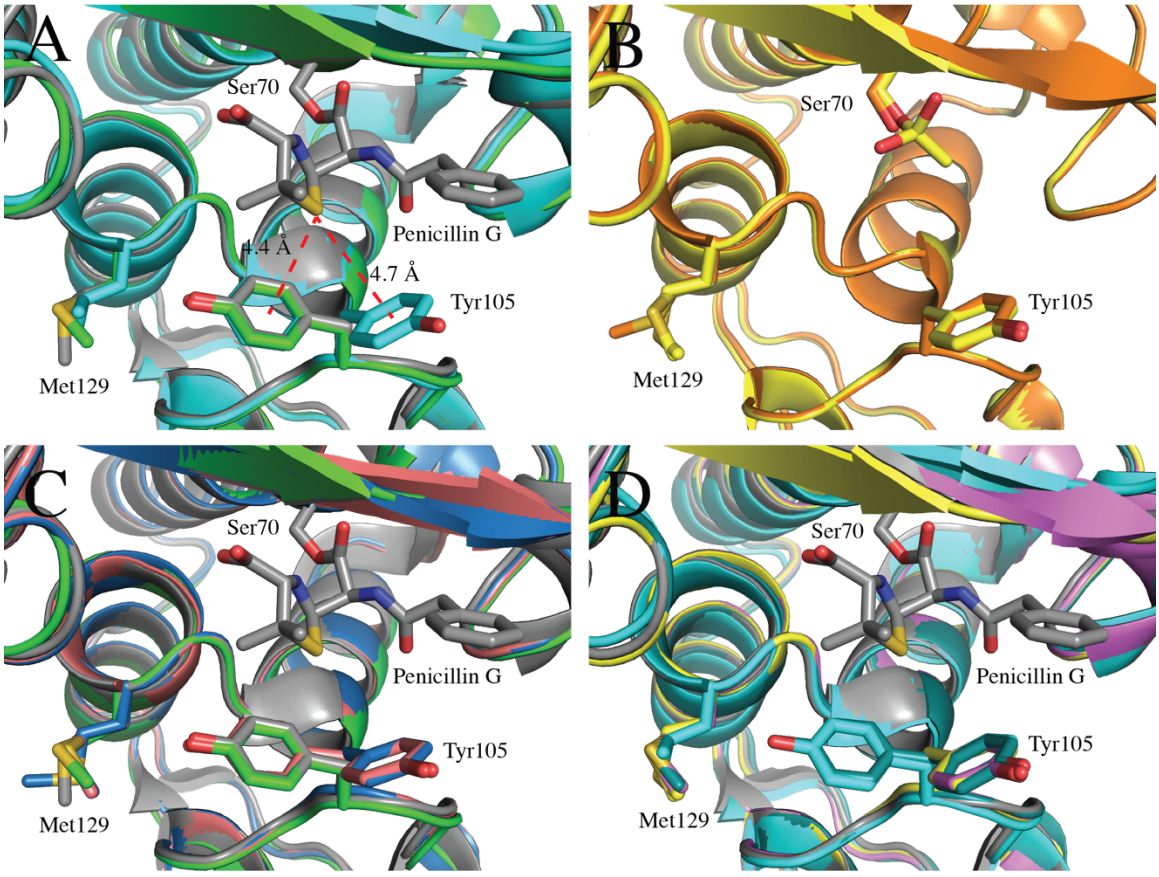


Figure 3.4.3.1 Models of all the crystallized TEM-1 variants at pH 4.9 superimposed with wild type TEM-1 (PDB ID: 1btl) in green, GNCA (PDB ID: 4b88) in cyan and model of TEM-1 covalently bound to penicillin G (PDB ID: 1fqg) in grey. The conformation of Tyr105 varies in the models and no pattern between function against ampicillin and Tyr105 conformation is observed. (A) In wild type TEM-1, (green and grey) the phenol of Tyr105 points toward Met129 and makes a sidechain-backbone hydrogen bond with Met129. In GNCA, Tyr105 adopts the "outward" conformation with the phenol pointing toward solvent. The distance between the sulfur atom on the thiazolidine ring of penicillin G Ser70 adduct and the center of the Tyr105 aromatic ring is 4.4 Å when Tyr105 is in the "inward" conformation. In the "outward" conformation, the sulfur atom on the thiazolidine ring is 4.7 Å. (B) Model of Flx226b at pH 4.9 (yellow) with a Ser70 adduct overlayed with Flx226b at pH 7.5. Tyr105 is in the same "outward" conformation. (C) Models of apo TEM-1 PDB ID: 1btl (green), holo TEM-1 PDB ID: 1fqg (grey), Flx226a (salmon), Flx226c (blue) overlayed. Tyr105 is observed in an "outward" conformation in Flx226c and in two states in Flx226a. (D) Models of GNCA (cyan) and Rdg44c (teal) with Tyr105 in two states. Tyr105 is observed in the "outward facing conformation in models of Flx22b (yellow) and Rdg262b (violet).

Initially, I hypothesized that the "outward" conformation of Tyr105 was indicative of reduced function against penicillin derivatives. This hypothesis originated from the observation that Tyr105 was observed in the "inward" conformation in native TEM-1 (PDB ID: 1btl) while in the inhibited structure, (PDB ID: 1blh), Tyr105 is in the "outward" conformation. In addition, in the model of GNCA, the ancestral homologue of TEM-1 with reduced functionality against ampicillin and increased functionality against cephalosporins, Tyr105 is observed in the two conformations. The reduced activity of GNCA could be explained by the increased flexibility of Tyr105. However, the models of the TEM-1 variants from this study indicate that Tyr105 conformations are highly variable in the apo enzyme and are not linked to changes in function against penicillin-derived antibiotics. The same Tyr105 dual conformation observed in models of GNCA is also observed in the Rdg44c, and the Flx226a models. Rdg44c has a MIC_{amp} value of 26 μg/mL, which is lower than GNCA with a MIC_{amp} value of 43 μg/mL; while Flx226a has a MIC_{amp} value on par with wild type TEM-1 (1500 μg/mL). In the Flx226b, Flx226c and Rdg262b models, Tyr105 adopts the outward-facing conformation only. Flx226b, Flx226c, and Rdg262b have MIC_{amp} values of 376 μg/mL, 1500 μg/mL and 26 μg/mL, respectively (Table 1). In none of the models was Tyr105 observed solely in the inward-facing conformation. Therefore, the hypothesis that the Tyr105 conformation is indicative of native function is unsupported with these data.

3.5 Conclusions

Crystallization studies of the TEM-1 variants with varying abilities to hydrolyze ampicillin were performed with the hopes that they would provide a structural

explanation for the observed changes in function and thermostability. Flx226a, Flx226b, Flx226c, Rdg44c, Rdg262b were crystallized, and models were analyzed for changes in structure with respect to wild type TEM-1. Initial analysis focused on the active site and on the large amount of electron density surrounding catalytic Ser70. It was concluded after data was collected on a second crystal of Flx226b at a higher pH that the electron density was due a pH dependent acetate adduct on Ser70. Aside from the presence of an acetate adduct, in all the variants, the catalytic residues in the active site, Lys73, Ser130, Glu166, and Lys234 are in the same conformations as wild type TEM-1.

In the models, Tyr105, a conserved residue that makes up one of the edges of the active site, was observed in two different conformations, an “inward” and “outward” conformation. It was initially hypothesized that the “outward” conformation could have been indicative of reduced functionality. However, this does not seem to be the case as there was no observable pattern between enzyme function and Tyr105 conformation.

3.6 Acknowledgements

This research was funded by The National Science Foundation, grant number 1901709. I thank Dr. J. Nathan Henderson for his contributions toward this work in the form of data collection and in teaching me the ins-and-outs of crystallography.

Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No, DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health,

National Institute of General Medical Sciences (P30GM133894). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

CHAPTER 4

REMODELING THE LUCIFERASE ACTIVE SITE TO ACCOMODATE NOVEL LUCIFERIN ANALOGUES

This chapter is adapted from the following publication: "Love A., Caldwell, D. R., Kolbaba-Kartchner, B., Townsend, K. M., Halbers, L. P., Yao, Z., Brennan, C. K., Ivanic, J., Hadjian, T., Mills, J. H., Schnermann, M. J., & Prescher, J. A. (2023). Red-Shifted Coumarin Luciferins for Improved Bioluminescence Imaging. Journal of the American Chemical Society, 145(6), 3335–3345.

This chapter is also adapted from a publication currently in review. "Expedient Synthesis and Characterization of π-Extended Luciferins" by Donald R. Caldwell, Katherine M. Townsend, Bethany Kolbaba-Kartchner, Tanya Hadjian, Joseph Ivanic, Jeremy Mills, Jennifer A. Prescher, Martin J. Schnermann.

In both publications, Bethany Kolbaba-Kartchner performed the Rosetta design methods and analysis.

5.1 Abstract

In vivo multicomponent bioluminescence imaging requires a set of robust luciferase-luciferin pairs that can emit in the near-infrared range. However, novel luciferin analogues with red-shifted emission spectra often have a larger chemical footprint than the native D-luciferin that emits in the yellow-green range and the luciferase active site must be re-engineered to accommodate them. A common approach to mold an active site to accept a new ligand is to create a library of variants and to screen or select them for the desired function. The challenge lies in how to identify locations in the protein to target for these libraries. To this end, a high-throughput computational

platform was developed. This platform consists of two parts: a Rosetta docking and design combined with a data analysis portion that uses Jupyter notebooks. With this platform, five novel luciferin-luciferase systems were identified with a λ_{max} of > 650 nm making them good tools for multicomponent imaging techniques.

Analyzing a case study of one library campaign to identify luciferase variants that can accommodate a series of luciferin analogues inspired by the coumarin molecule indicates the successfulness of this platform. All five of the residues that were identified through library screening as being important to processing the coumarin-derived luciferin analogues were identified in the RosettaDesign platform. This work showcases a new platform for identifying locations to target for library design.

5.2 Introduction

Fireflies, potentially the most nostalgic insect, are often associated with fond memories of warm summer evenings as children delight in catching them in jars. Members of the Lampyridae family, fireflies attract curious children thanks to a bioluminescent 62 kDa enzyme aptly named luciferase (Fluc) that emits bursts of yellow-green light from their abdomen to communicate, attract mates, and deter predators (Kaskova et al., 2016). On an atomic level, in an aerobic environment, luciferase catalyzes the oxidation of a small molecule, D-luciferin, resulting in an adenylate intermediate. Oxidation of the adenylated intermediate leads to an electronically excited state oxyluciferin. As the excited state relaxes to its ground state, it releases energy in the form of a photon of light with a λ_{max} of 554 nm(Branchini et al., 2005) (Figure 5.2.1).

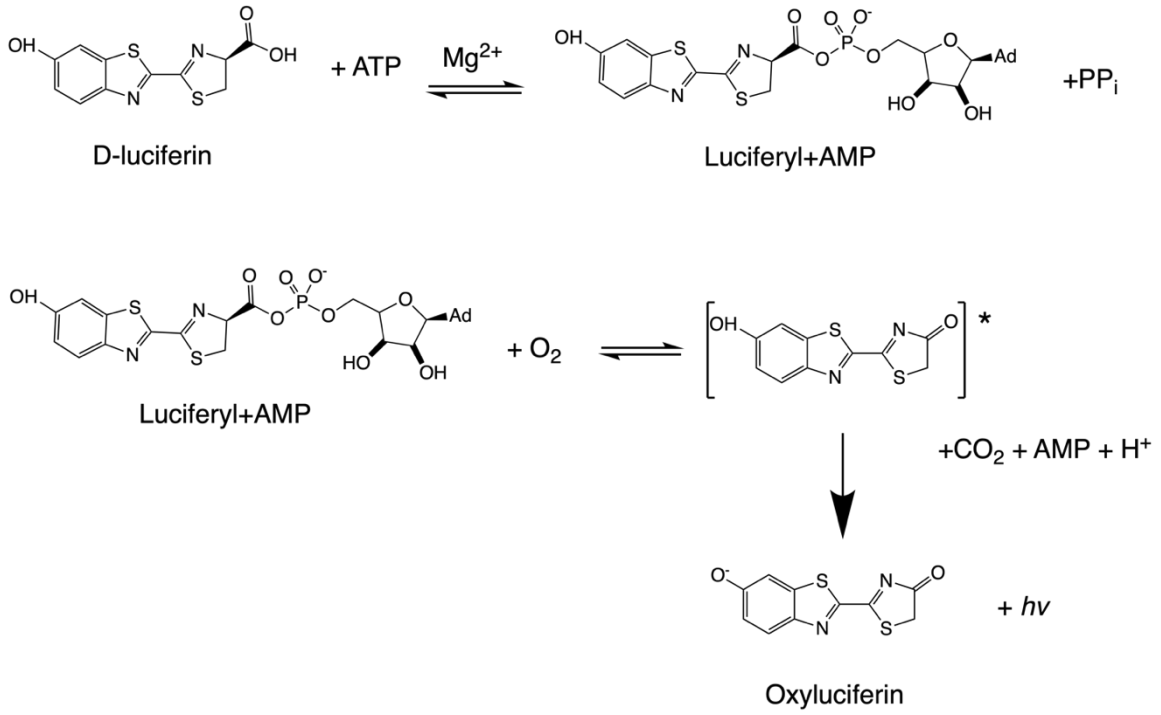


Figure 4.2.1. The luciferase chemical reaction is composed of two half reactions resulting in the emission of a photon of light. The natural substrate, D-luciferin, is first adenylated forming a luciferyl-adenylate molecule. The luciferyl-adenylate molecule is subsequently oxidized producing an electronically excited oxyluciferin that releases a photon of light upon relaxation.

Because the bioluminescent reaction does not require an external source of light, luciferase can not only be used to create a lantern out of an old pickle jar, but also, and (potentially) more importantly, to noninvasively visualize events in the cell like protein transport, protein-protein interactions, cell proliferation, and gene expression. However, a disadvantage of using native firefly luciferase for cell-based, and more broadly, tissue-based applications, is that molecules endogenous to tissues like hemoglobin and melanin absorb in the same region as the λ_{max} of Fluc; thereby reducing the available signal. This

is especially problematic when attempting to visualize events in deeper tissues (Rice et al., 2001).

Red-shifting the reaction emission maximum from 554 nm to > 650 nm would better isolate the luciferase signal from endogenous chromophores and increase the signal-to-noise ratio for better resolution. In addition, the development of new luciferase/luciferin pairs with varying emission maxima would confer the ability to visualize multiple cellular events simultaneously providing a more comprehensive view of complex biological phenomena. Thus, work has been undergoing to develop luciferase-luciferin pairs with varying, but distinct longer-wavelength emission spectra (Love et al., 2023; Yao et al., 2018). This has been successfully done using Directed Evolution techniques (F. H. Arnold et al., 2001; Renata et al., 2015) notably in the case of the development of the Akaluc/Alakumine luciferase/luciferin pair (Iwano et al., 2018). With 21 rounds of random mutagenesis, Iwano et al. identified a luciferase with 28 mutations relative to native Fluc with a λ_{max} of 650 nm. Despite the success of this method, when one is initially agnostic with respect to locations to mutate in a library, the many rounds of random mutagenesis required is very time and resource intensive.

Rational design offers an alternative approach to identifying target mutations for library design. Typically, identifying locations to target is done by mining the literature for potential sites, performing alanine scans, or by structural studies of protein crystal structures. Although these methods have worked well in the past, they are low throughput. In addition, even when a suitable crystal structure of a protein bound with its endogenous ligand exists, it is still challenging to identify how a novel ligand will bind in the active site and to translate that to identify residues to target in a library.

To this end, I worked on developing a high-throughput computational platform to identify target locations in the active site to include in a semi-rational Combinatorial Codon Mutagenesis (CCM) library (Belsare et al., 2017) aimed at identifying novel luciferin/luciferase pairs with a λ_{max} > 650 nm. The platform was composed of two parts: docking/design and analysis. The first part employed the RosettaMatch algorithm to dock novel luciferin analogues into a starting scaffold and the RosettaDesign algorithm to identify target locations for an experimental library (Richter et al., 2011). The second part employed Jupyter notebooks to analyze Rosetta output to 1) identify holo-enzymes where the ligand was in a conformation believed to be conducive to a productive chemical reaction and 2) calculate the number of times a particular location in the active site was mutated to accommodate the novel luciferin. Because the downstream application of this data was an CCM library, targeting "hot spot" residues identified by Rosetta, the data of interest was less the identity of the mutation, but the locations where Rosetta predicted mutations were necessary to accommodate the ligand in the active site. This workflow resulted in the development of five (CouLuc-3-NMe_2, two FPLucs, CouLuc-1-NMe_2 and -OH) luciferase-luciferin pairs with emission maxima > 600 nm (Figure 4.3.2.1).

4.3 Materials and Methods

To develop more red-shifted luciferase-luciferin pairs, our collaborators in the Martin Schnermann lab at the NIH synthesized sets of molecules based on three small molecule fluorophores: coumarin, 1-naphthol, and the red fluorescent protein chromophore (Figure 4.3.2.1). They also developed more extended luciferin analogues based on the Akalumine structure (Iwano et al., 2018). Individual molecules in each set differed by the length of the pi wire and the addition of functional groups on the

heterocycles (Figure 4.3.2.1). These additions served to extend the conjugation of the molecule thereby increasing the λ_{max} of the emission spectra. These novel luciferin analogues were not compatible with native Fluc as the emission max was typically very low at the outset. Therefore, our collaborators in the Jennifer Prescher lab at the University of California, Irvine set up CCM libraries using the NNK degenerate codon to screen for Fluc variants that could process the luciferin analogues with greater efficiency. To reduce the number of rounds of Directed Evolution they approached the Mills lab to employ Rosetta to identify target locations for library design.

Identifying target locations for a semi-rational library design began with building a solid model system which consisted of two separate, but equally important parts: the ligand and the scaffold. Unsurprisingly, both structures required preparative steps prior to modeling.

4.3.1 Preparation of the Luciferase Scaffold. In solution, luciferase folds into two domains: a large N-terminal domain and a small C-terminal domain (Conti et al., 1996). There is a dynamic cleft between the two domains that closes upon ligand binding undergoing a massive change in protein conformation including a 25 Å displacement of the C-terminus (Figure 4.3.1.1.A). Additionally, the active site itself undergoes drastic conformational changes upon ligand binding with the loop composed of residues 315-320 moving 3.3 Å to open up the space (Figure 4.3.1.1.B). With a dynamic enzyme like luciferase, it was imperative to choose a structurally accurate input for modeling. Ideally, a starting structure is one where the conformation of the protein is conducive to ligand binding. Fortunately, such a structure existed in Fluc co-crystallized with luciferyl-AMP (DLSA), a high-energy intermediate analogue (PDB ID: 4g36) (Sundlov et al.,

2012).Other crystal structures of Fluc or its homologues crystallized without a ligand, were in protein conformations inconducive to binding.

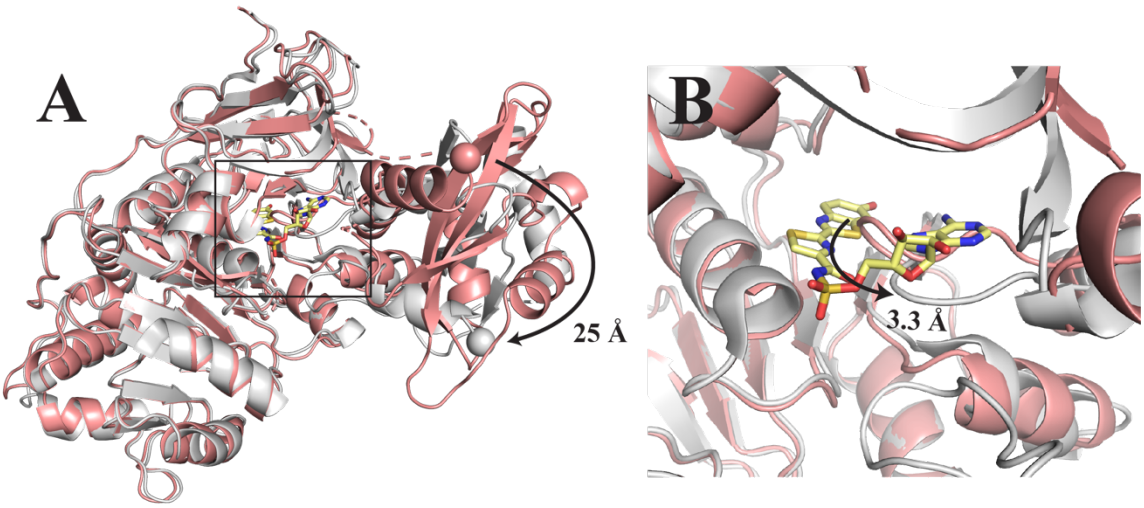


Figure 4.3.1.1 Model of the dynamic Fluc. (PDB ID: 4g36) (grey) with bound high energy intermediate DLSA (yellow sticks) and Fluc (PDB ID: 1lci) (salmon) without a bound ligand. (A) C-terminal residues are shown as spheres to illustrate the ~25 Å domain movement upon ligand binding. (B) This domain movement causes the loop composed of residues 315-320 to move 3.3 Å, opening up the active site for productive ligand binding.

Initial attempts to prepare the Fluc scaffold for modeling followed a well-established protocol to remove all non-proteinaceous molecules prior to equilibrating the scaffold with the Rosetta score function. However, close inspection of the equilibrated models overlayed with the input crystal structure showed that without a ligand in the active site, loop 315-320 adopted the lower energy conformation observed in the unliganded crystal structures. This “closed conformation” of the loop precluded productive ligand binding. Therefore, preparation of the structure included the co-crystallized ligand in the active

site which constrained loop 315-320 to the open conformation. This small change was integral to all downstream applications.

Our collaborators in the Jennifer Prescher lab have worked extensively with luciferase and over the years they have developed libraries of starting scaffolds for initial hit generation. Often, to get their foot-in-the-door they begin their selection efforts by screening for the best starting scaffold from this set. The majority of the starting scaffolds were variants of Fluc and contained from 2-28 mutations to wild type. Initially, to prepare models of the starting luciferase variants, I made the mutations in the model scaffold using an input "resfile" as an initial step prior to sampling the binding mode of the novel ligand. However, as I optimized the platform, I changed the order of events so that the mutations were made during the design step. The reasoning behind this change was two-fold. First, preparing the Fluc scaffold once for docking and design cut out a step and made the platform less computationally time-intensive. Second, I reasoned that making the required mutations in the context of a bound ligand would provide a better view of the interplay between the scaffold and the ligand.

4.3.2 Preparation of the Luciferin Analogues. The second preparatory step was to generate the ligand model. This step went through quite a few iterations before I settled on the best method because there were numerous parameters that went into this decision. As previously mentioned, the luciferase chemical reaction involves two half reactions: adenylation of the luciferin followed by oxidation (Figure 4.2.1). The first decision on how to properly model the novel luciferin analogues was whether the luciferin should be modeled as its starting structure or as the reaction intermediate, an adenylated luciferin. Because the product molecule is structurally very similar to the starting molecule, it was

not a contender in this decision. Since current models of enzyme catalysis assert that the active site stabilizes the transition state intermediate. Therefore, modeling the luciferin analogue as a reaction intermediate was the highly favored option. Nevertheless, the starting structure of the luciferin analogues had a smaller chemical footprint and it was reasoned that it might be easier to dock the starting molecule as a first pass.

Computational experiments were performed where the luciferin analogue was docked as the starting structure. This approach was quickly abandoned because of the flexibility of the molecule. It was difficult to obtain docked models where the 4-carboxy-thiazoline ring of the analogue was in the native geometric orientation in the active site as identified by a structural study of the 4-carboxy-thiazoline ring on DLSA in the crystal structure. Therefore, the luciferin analogues were modeled as the intermediate adenylated luciferin form. To do this, the adenylated moiety of the DLSA structure from the co-crystallized Fluc model (Figure 4.3.2.1.D) was conjugated to the luciferin analogue to create a model of the luciferin intermediate. The adenylated moiety of the DLSA model thus served as a handle to anchor the luciferin analogues in the active site (Branchini et al., 2005).

The ligands were built in Avogadro an open-source molecular builder and visualization tool. Version 1.2.0. http://avogadro.cc/ using the crystallized adenylated moiety of the DLSA model as the base (Hanwell et al., 2012). They were energy minimized using the UFF force field (Rappe et al., 1992) because it is a good general energy minimization algorithm. Another force field available in the Avogadro software, the Merck Molecular Force Field (MMFF94), was employed experimentally, but was found to be less than optimal. MMFF94 is a force field built to model a molecule in an aqueous environment (Halgren, 1996).Therefore, intramolecular interactions in the

molecule are upweighted. This caused the luciferin to take on a more compact structure during the energy minimization where interactions between the adenyl moiety and the heterocycles at the end of the pi wire are stronger. Post energy-minimization, the pi wire was more curved and no longer linear. Consequently, the luciferins subjected to the MMFF94 force field would not fit in the luciferase active site. I reasoned that because the luciferyl-AMP intermediate was only present in the protein active site environment, the ligand was less likely to be in the same geometric conformation as the MMFF94 predicted. The UFF force field minimization maintained the extended pi wire conformation that was more conducive to ligand binding. In all energy minimization events, the adenyl moiety, or the "handle" was always constrained from moving.

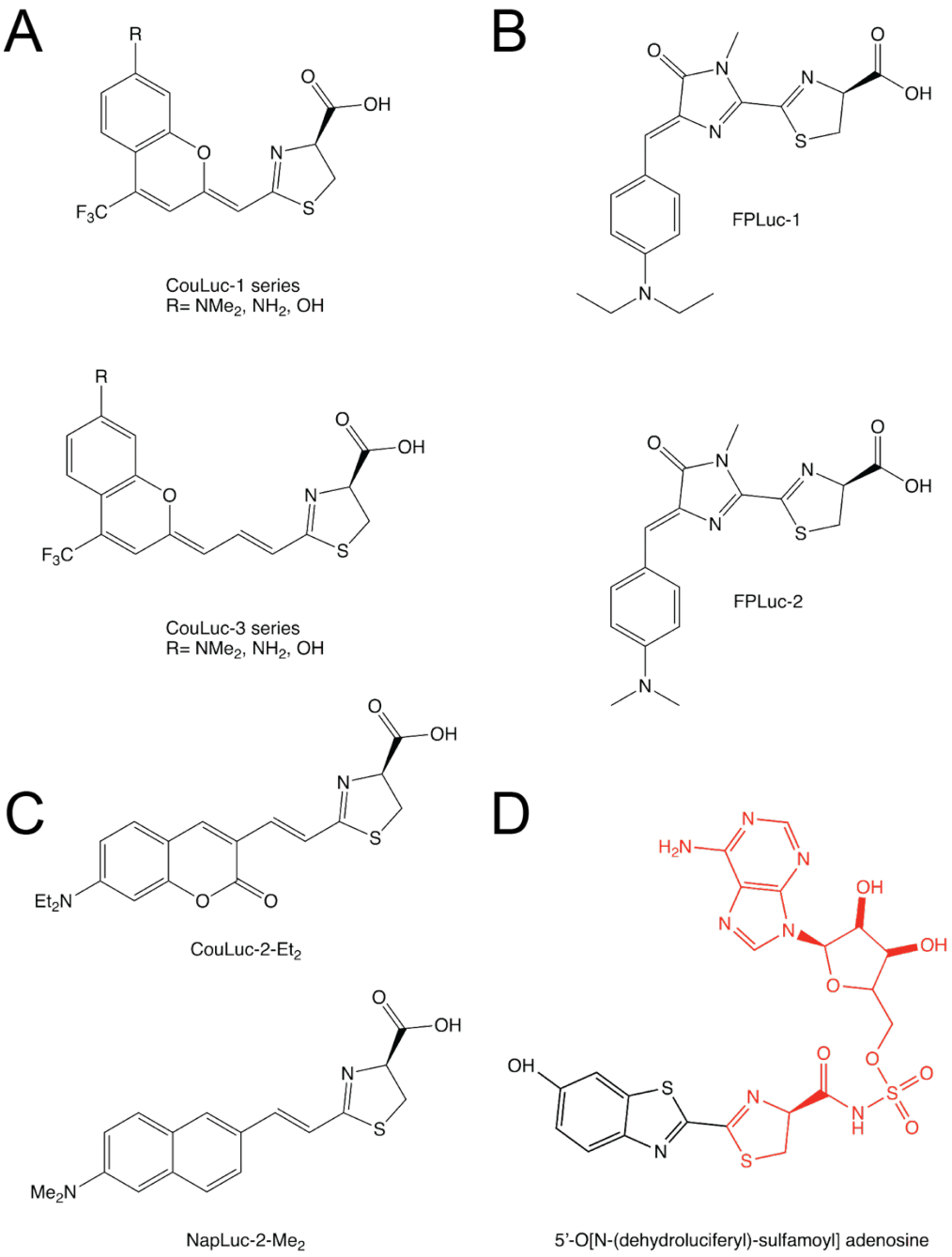


Figure 4.3.2.1 Luciferin analogues that were incorporated into the rationally designed library protocol. (A) the CouLuc series built on the coumarin moiety. (B) the FPLuc series built on the red fluorescent protein chromophore (C) the NapLuc and CouLuc-2 series built around naphthalene and coumarin, respectively. (D) 5'-O[N-(dehydroluciferyl)-sulfamoyl] adenosine (DLSA), an analogue of D-luciferin that was co-crystallized with Fluc (PDB ID: 4g36) and used as the base for building the luciferin

analogues. The red portion is the adenyl moiety that was added to the novel luciferin analogues for computational modeling.

Originally, I performed a rigid docking of the luciferin analogues into the active site; relying on the Rosetta software to sample limited ligand torsion angles. However, I quickly discovered that this method was not conducive to identifying potential binding modes. When experimental data informed me that the one of the luciferins that I had previously believed impossible to dock into the Fluc scaffold generated light, I knew that my method needed to be improved. I increased the ligand sampling space by incorporating a file of enumerated ligand conformations generated using the OpenEye Omega software (Hawkins et al., 2010a). To limit the number of luciferin conformers in the library to only those that would be productive, I constrained the torsion angles in the handle of the ligand composed of the adenyl moiety to the carbon past the thiazoline ring to the angles calculated from the crystal structure. Only the portion of the luciferin analogue that was composed of the starting molecule was allowed to sample torsional space (Figure 4.3.2.2). The inclusion of a ligand conformer library was integral to identifying potential binding modes of the analogues.

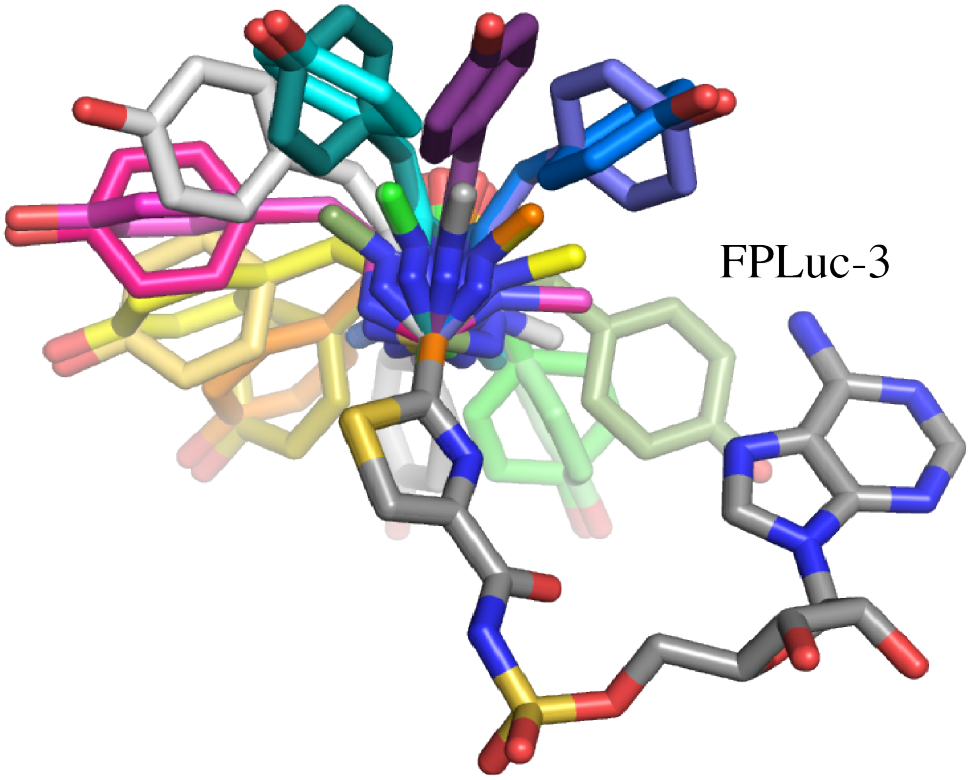


Figure 4.3.2.2 Sample conformer library for FPLuc-3. For conformer generation, the torsion angles of the adenyl moiety (modeled as grey sticks) were constrained to the observed angles in the Fluc crystal structure (PDB ID: 4g36). The portion making up the starting molecule near the thiazoline ring were allowed to sample torsion space (shown as colorful sticks).

4.3.3 Docking the Ligands and Sculpting the Luciferase Active Site. To determine the native-like geometric alignment of D-luciferin in Fluc, a structural analysis of two crystal structures was performed. The first was one of a luciferase from the Japanese Genji-botaru (Luciola cruciate), a homologue of Fluc, co-crystallized with products AMP and oxyluciferin (PDB ID: 2d1r) (Nakatsu et al., 2006). The second was of Fluc co-crystallized with a luciferyl-AMP intermediate analogue, DLSA, (PDB ID: 4g36) (Sundlov et al., 2012). It was discovered that the carbonyl proximal to the thiazoline ring that reacts with ATP to form the luciferyl-AMP intermediate was within

3.4 Å of Cα G315 on the 315-320 loop. Therefore, the Cα of G315 was used as an anchor point for docking the luciferins in a native-like geometry. This measurement was incorporated into the constraints imposed on the RosettaMatch algorithm to properly dock the novel luciferins into the Fluc scaffold.

The RosettaMatch algorithm was employed to dock each novel luciferin in a native-like orientation in the prepared native Fluc structure (PDB ID: 4g36) (Tinberg &Khare 2017; Zanghellini et al., 2006). Docking was done in rounds with increasing weight placed on the constraints on ligand placement. The first round typically contained loose constraints to verify that the ligand could semi-productively bind in the active site. In subsequent rounds, the constraints were tightened until RosettaMatch could no longer find binding orientations for the luciferin. Docked models were chosen for design based on the closest adherence to the native binding mode as defined by 1) the distance of the carbonyl to loop 315-320 and 2) the angle between the Cα of G315 and the carbonyl and proximal atoms on the luciferin analogue as calculated from the crystal structure.

Output files from the docking step were used as input to the RosettaDesign algorithm. RosettaDesign was used to randomly mutate residues near the docked luciferin analogue. The viability of each mutation was assessed using the Rosetta score function (Alford et al., 2017). Mutations that alleviated clashes between ligand atoms and protein atoms or that introduced new interactions (e.g., hydrogen bonds or hydrophobic packing interactions) between the ligand and protein typically improved the Rosetta energy. In this way, Rosetta was used to restructure the active site to accommodate novel ligands that initially were structural misfits.

4.3.4 Analyzing the Output Models. Extracting data from the RosettaDesign outputs was where the most interesting platform development occurred. Rosetta proper is equipped with built-in methods to analyze potential output during the design process itself. These methods, called "filters," are typically a robust way to eliminate models that do not meet specific criteria (Fleishman et al., 2011).The benefit of filters is that they eliminate all models that do not pass certain criteria on the fly, so they are never outputted. This is desirable when an appropriate filter exists, but in this case, one did not. Thus, the analysis of the Rosetta designs took place post-design.

The data that I wanted to extract from the RosettaDesign output was what locations did Rosetta mutate and how often that particular location was chosen in the context of ligand positioning in the active site. Since this analysis was integral to identifying initial target locations for a NNK library, the residue identity of the mutation was unimportant. The output pdb files from RosettaDesign were analyzed using code collected into Jupyter notebooks increasing both the reproducibility of the analysis and allowing it to be high throughput.

Each docked input structure, of which there were typically 10-15, were very similar with only slight variations in the torsion angles and geometric positioning of the ligand gave rise to ~10-25 output designs. Often the ligand placement in the output designs was similar to the input structure, but this was not a hard-and-fast rule. Rosetta typically performed rigid-body ligand translations during the design and side chain repacking steps. The variability of the Rosetta Monte Carlo algorithm made it a challenge to analyze the output designs in a high-throughput way. This challenge was overcome by binning the output files into sets based on observed ligand conformation in the active site.

This ligand-centric approach allowed for a clearer view of how the different ligand conformations affected the mutation locations chosen by Rosetta.

To bin the models, the distance was calculated between an atom on the heterocycle of the luciferin analogue and an Cα on a residue in the active site in proximity of the heterocycle atom. (Figure 4.3.4.1)

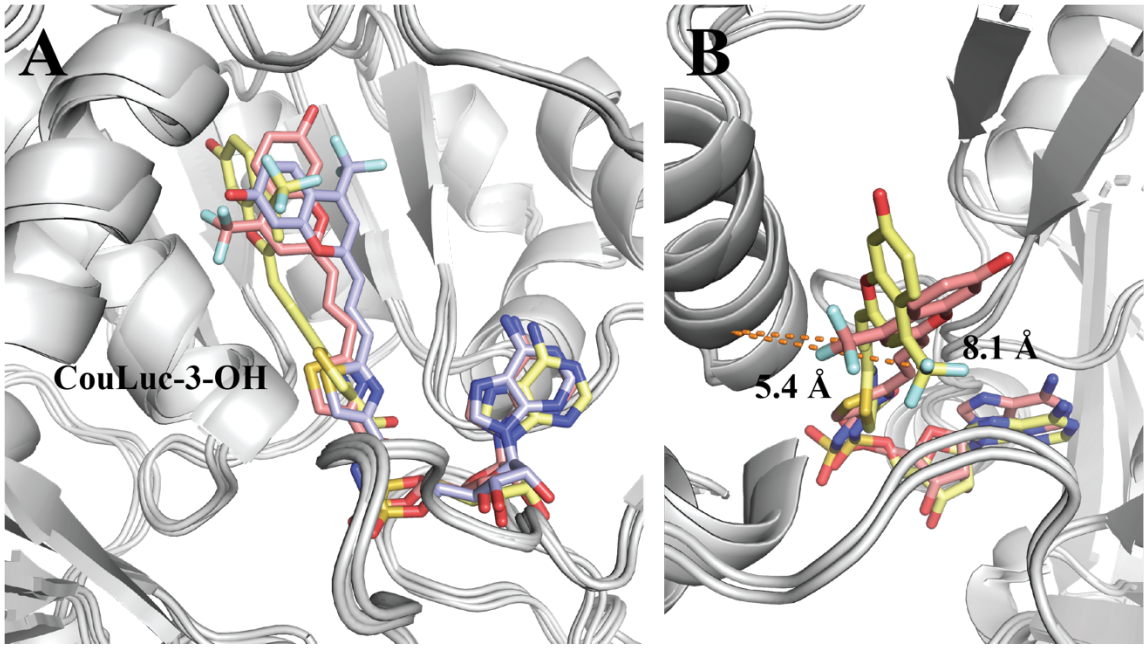


Figure 4.3.4.1 Select Rosetta models of CouLuc-3-OH bound in the active site of Fluc. (A) CouLuc-3-OH adopts multiple conformations in the active site (B) To identify how the conformation of the luciferin analogue affects the locations mutated by Rosetta, the models were binned by the distance between a backbone atom in the active site and an atom on the ligand. There are two bins represented in this example: 5 Å and 8 Å.

The calculated distance was rounded up to the nearest integer and that integer was the bin. Typically, one RosettaDesign run contained from 4-8 bins with ~1-50 files in each bin. Within each bin, the locations where Rosetta made a mutation (relative to the input scaffold sequence) were identified and the frequency of each mutation location was

calculated. In addition, the frequency of each mutation location was also calculated for all the models regardless of bin status. This analysis provided a global view of the number of mutations as well as a very specific view by ligand binding mode.

4.4 Results and Discussion

4.4.1 The CouLuc-3 Series as a Case Study Five novel luciferase-luciferin pairs with robust red-shifted photon output were identified using the data generated from the luciferase platform. To assess the efficacy of the platform, the CouLuc-3 series of luciferin analogues can be used as a case study. The campaign to identify a Fluc variant that would accommodate the coumarin derived luciferins (Figure 4.3.2.1.A) was particularly robust and involved iterative rounds of design-experimentation. The CouLuc-3 series includes three luciferins that differed in the electron-donating groups (-NMe_2, -NH_2, and -OH) installed on the coumarin heterocycle. These groups differ significantly in size from nine atoms in -NMe_2 to three atoms in NH_2 and two in -OH. When the light output of the three luciferins was measured with native Fluc, photon output of CouLuc-3-NMe_2 and CouLuc-3-NH_2 was too dim to be measured. However, CouLuc-3-OH was found to have the brightest emission maximum and a λ_{max} of 730 nm. I was reasoned that CouLuc-3-OH, with its smaller chemical footprint was better accommodated in the active site than the larger analogues in the series. However, the starting point of the emission maximum of CouLuc-3-OH was insufficient for imaging applications. Experimental efforts focused on developing Fluc variants for CouLuc-3-OH and CouLuc-3-NMe_2 because these two molecules were successfully synthesized in high yields while CouLuc-3-NH_2 was not. Additionally, it was reasoned that Fluc variants that could process the

smaller CouLuc-3-OH or the larger CouLuc-3-NMe_2 would likely be able to process CouLuc-3-NH_2 sufficiently well for imaging applications.

To engineer more robust luciferase/CouLuc-3-OH and -NMe_2 pairs, the computational luciferase platform was employed iteratively with input from experimental data. To identify target locations for the first generation RosettaDesign library, the CouLuc-3 luciferases were docked into wild type Fluc. Based on the docking output, it became immediately apparent that there was a significant steric clash between the sidechain of R218 and both luciferins because every design included a mutation to a smaller residue (Figure 4.4.1.1). When Rosetta was prohibited from mutating R218, it adopted a high-energy conformation. Therefore, this location was included in the first generation RosettaDesign library. Along with R218, there were 40 more locations that were identified as potential targets in the computational analysis. Most of these locations were in the active site, but there were also a few mutations in the AMP-binding domain. With respect to target locations in the AMP-binding domain, it was discovered after five rounds CCM libraries that none of the targets in the AMP binding region led to any improvements in CouLuc-3 processing. Therefore, future campaigns omitted this region in library design.

Our collaborators chose 20 locations to target for the RosettaDesign generation 1 library (Figure 4.4.1.A).

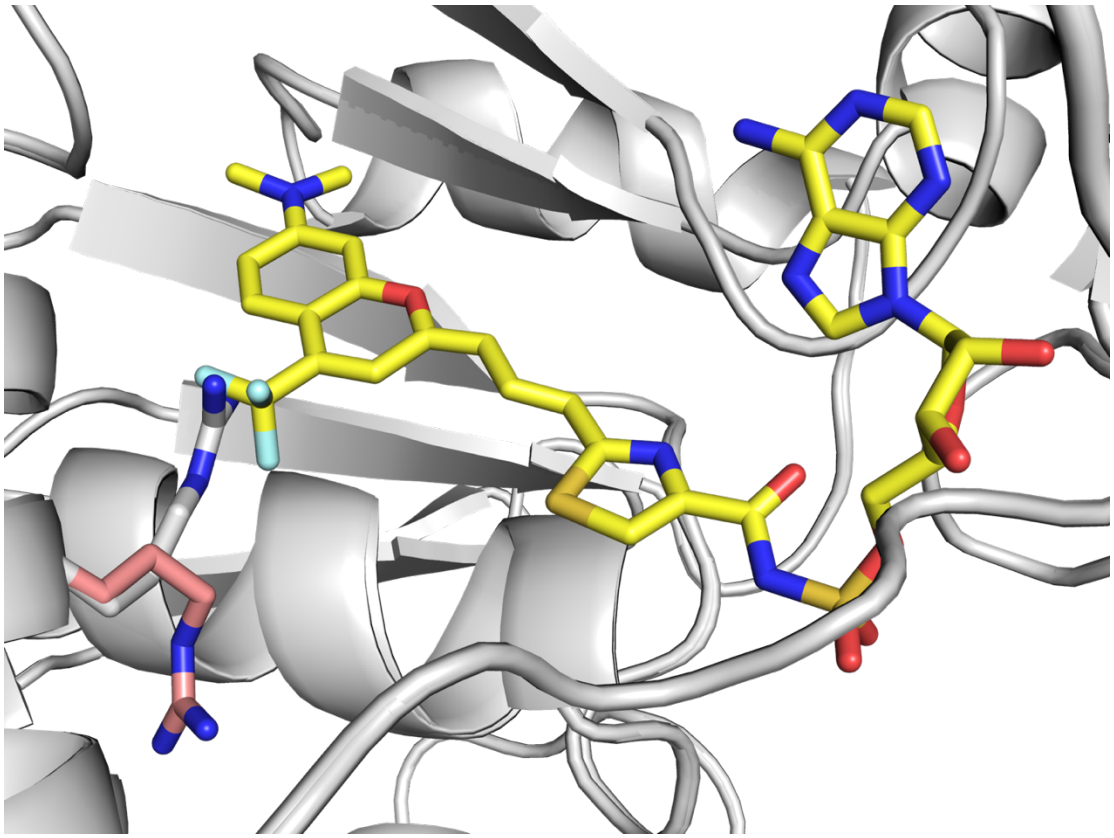
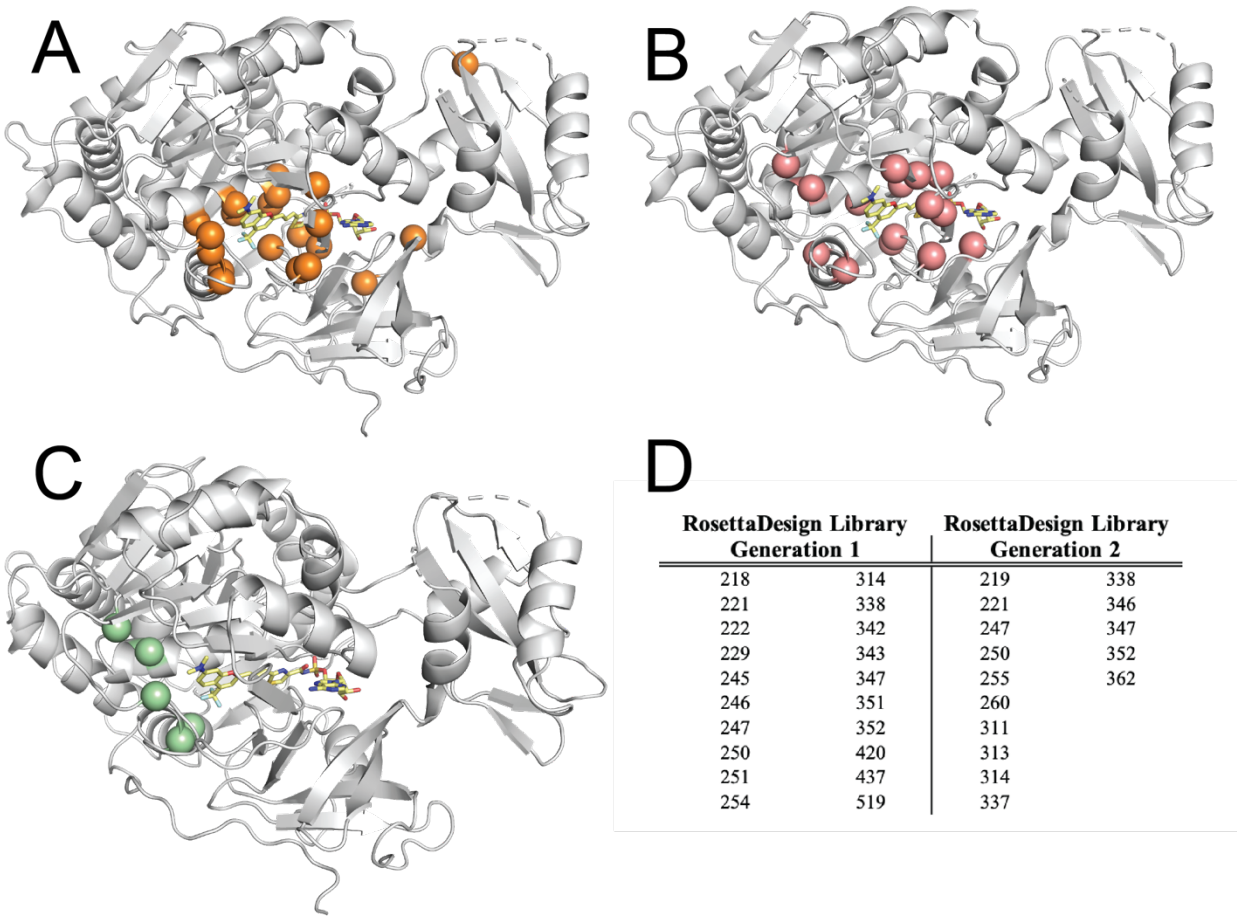


Figure 4.4.1.1 A model of CouLuc-3-NMe_2 bound in the Fluc active site. R218 (grey) is in the conformation modeled in the crystal structure of PDB ID: 4g36. R218 (salmon) adopts a high energy conformation when CouLuc-3-NMe_2 is bound.

The brightest "hits" from the first generation RosettaDesign library contained the following mutations: R218S/V and N229F. However, the photon output was not as robust as desired, therefore a second round of modeling was undertaken using the hits from the first generation library as the starting scaffolds. Three scaffolds were prepared: R218S alone; R218S with N229F; and R218V with N229F. The second round of target hunting identified another ~ 30 potential target locations. Most of the locations were also found in the first set. However, a few new locations including two that would end up being important to CouLuc-3 processing appeared in low frequency. The first important location was F260 which was mutated 5% of the time when CouLuc3-NMe_2 was bound

and 2% of the time when CouLuc3-OH was bound. The second was Y255 which was mutated in 14% of the designs regardless of luciferin analogue bound. These locations along with 7 locations from the first-generation library and six new locations made up the second generation RosettaDesign library (Figure 4.4.1.2.B). Top hits from this library identified H221, Y255 and F260 as being important to CouLuc3 processing.



RosettaDesign Library Generation 1		RosettaDesign Library Generation 2	
218	314	219	338
221	338	221	346
222	342	247	347
229	343	250	352
245	347	255	362
246	351	260	
247	352	311	
250	420	313	
251	437	314	
254	519	337	

Figure 4.4.1.2. Locations chosen for the RosettaDesign libraries. (A) Locations chosen for the first generation (B) Locations chosen for the second generation (C) Locations that were identified as "hits" and (D) a table of the mutations chosen.

The first generation library included three hits that were focused on mutations at the following locations: R218, N229 and H221, all of which were Rosetta "hot spots."

The second generation RosettaDesign library identified two more target locations: F260 and Y255. In an effort to increase emission output, two more error-prone PCR and random mutagenesis libraries were designed and screened. However, they failed to produce luciferase variants with emission output greater than those identified in the two RosettaDesign libraries.

The final output of the rounds of library screenings revealed two mutants with > 650 nm emission maxima for all three CouLuc-3 analogues and a total flux of 380 fold and 9 fold greater than wild type Fluc when processing CouLuc-3-NMe_2 and CouLuc-3-OH, respectively. The first mutant was named “Pistachio” after a Prescher lab tradition to name successful mutants after nuts. The second, was sadly neglected and kept its original library number of “709”. Pistachio contained the following mutations: R218S, H221I, N229F, and F260G. Unnamed mutant 709 contained the following mutations: R218V, N229F, Y255S, F260R. (Love et al., 2023).

Retrospectively analyzing the Rosetta-predicted hotspots with the mutations contained in the final mutants is an important step to improving the platform. Experimental data indicated that only five total mutations were necessary to process the CouLuc-3 analogues with sufficient photon output to be useful in biomedical applications. The mutation locations that were important to process all three luciferin analogues were R218, N229 and F260. One mutation was specific to the processing of CouLuc-3-NMe_2: H221. The mutation specific to CouLuc-3-OH processing was Y255. All five locations appeared in the Rosetta-guided library. Three of them, R218, H221, and N229 were predicted as targets in 100%, 73% and 54% of the models, respectively. Y255 and F260 were only predicted to be targets in 14% and 2% of the models,

respectively. However, Rosetta predicted 36 more “hot spots”. This large signal-to-noise ratio provides ample room for platform improvement that would likely be made possible with tighter constraints during the RosettaDesign step and on increasing the number of outputs from RosettaDesign.

4.4.2 Modeling Experimental Data with Rosetta Frequently, my collaborators would be baffled by an experimental result and I would be asked to create a model of the protein-ligand interaction to better understand the data. Of course, solving the crystal structure of the holo protein, or undertaking more experiments is the best way to develop hypothesis to explain experimental data, but I found that using Rosetta to create models provided context and was a useful way to generate hypothesis. There are three examples of opportunities where I used Rosetta to generate hypothesis for experimental data. In this section, I will go into detail on one of them. The other two examples can be found in the following publications from the Leconte lab and the Prescher lab (Love et al., 2023; Williams et al., 2023).

One of the questions that our collaborators wanted to answer was why the light output was so much more robust when NapLuc-2-NMe_2 was bound over when CouLuc-2-NEt_2 was bound in Akaluc. Akaluc is an engineered luciferase from the Miyawaki lab with 28 mutations from wild type Fluc, most of which are outside the active site (Iwano et al., 2018). Akaluc is one of the scaffolds that the Prescher lab routinely uses as a starting point for their engineering efforts.

Two luciferins that were synthesized by the Martin Schnermann lab: CouLuc-2-NEt_2 and NapLuc-2-NMe_2 were found to have differences in light output when processed by Akaluc even though they are structurally very similar. (Figure 4.3.2.1.C). In Akaluc,

CouLuc-2-NEt_2 had $10^{7.5}$ p/s total flux and NapLuc-2-NMe_2 had 10^{10} p/s total flux. We hypothesized that the increased steric bulk of the diethylamino substituent in CouLuc-2-NEt_2 (relative to the dimethylamino substituent in NapLuc-2-NMe_2), or the carbonyl on the pyrone ring of the coumarin moiety might have precluded productive binding of CouLuc-2-NEt_2 within the active site. To explore this, we first generated *in silico* models of NapLuc-2-NMe_2 and CouLuc-2-NEt_2, using Avogadro and then generated libraries of low-energy conformers of each of these substrates using OMEGA 4.2.1.2: OpenEye Scientific Software, Santa Fe, NM (Hanwell et al., 2012; Hawkins et al., 2010b). Because no structure of Akaluc has been reported, the conformer libraries were docked into a structure of Fluc (PDB ID: 4g36) using the RosettaMatch algorithm; user-defined constraints were employed in this docking step to ensure a native-like binding geometry was achieved. The Rosetta CoupledMoves algorithm was used to 1) introduce the 28 mutations present in Akaluc, 2) sample low-energy ligand conformations, 3) sample low-energy side chain conformations, and 4) sample backbone torsions to alleviate any clashes between the ligand and the protein (Ollikainen et al., 2015).The resulting models were then analyzed to assess whether differences in the binding modes of the two ligands were present. In our models, the bound CouLuc-2-NEt_2 substrate was observed in both the E and Z isomeric forms (Figure 4.4.2.1.A). When the Z isomer is present, the ligand is well accommodated within the active site; namely, the carbonyl on the pyrone ring is observed in a glycine-rich microenvironment in which the C_as of G315, G316, and G339 fall within ~3.4-6 Å of the carbonyl oxygen atom (Figure 4.4.2.1.B).

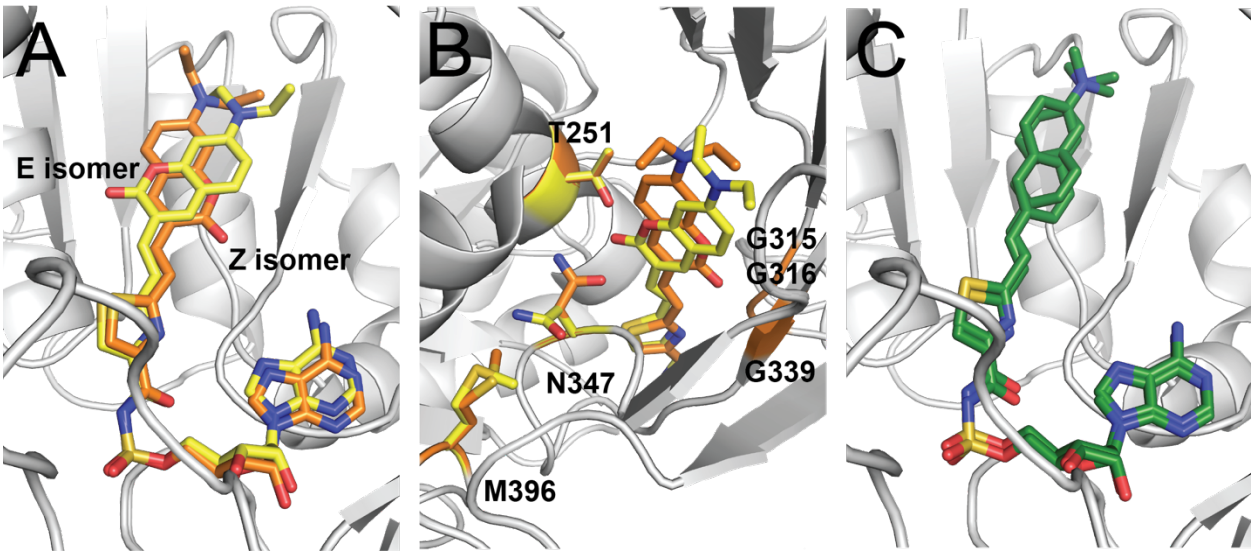


Figure 4.4.2.1 (A) The lowest scoring Rosetta models of E (yellow) and Z (orange) isomers of CouLuc-2-NEt_2 and bound in the Akaluc active site. (B) The Z isomer is well accommodated in the active site as the carbonyl on the pyrone ring is in a glycine-rich microenvironment with G315, G316 and G339 within 3.4-6 Å of the carbonyl oxygen atom. The CouLuc-2-NEt_2 E isomer is not well accommodated in the active site. In 70% of the models, the carbonyl on the pyrone ring is observed within 2.1 Å of Oγ of T251. The side chain of N347 must also rotate ~150° about the C_α and C_β bond toward the side chain of M396 to accommodate the ligand, forcing it into a high energy conformer. In 30% of the models, the distance between the O_δ of N347 and the C_ε of M396 is < 3 Å. (C) Two low scoring Rosetta models of NapLuc-2-NMe_2 bound in the Akaluc active site.

However, when the E isomer is bound, the carbonyl on the pyrone ring is not as well accommodated in the pocket. In 70% of the models, the carbonyl on CouLuc-2-NEt_2 is observed within 2.1 Å of Og of T251. In 30% of the models, the carbonyl is between 3-4.5 Å of the Og of T251, but the side chain of N347 must rotate ~150° about the C_a and C_b bond toward the side chain of M396 to accommodate the ligand. In 30% of the models with the E isomer bound, the distance between the O_d of N347 and the C_e of M396 is < 3 Å. In addition, the Rosetta scores of N347 and M396 are 5.3 and 0.47 REU, respectively which indicates that these residues are higher in energy and could preclude proper

binding of the ligand. The proximity of the carbonyl group on CouLuc-2-NEt_2 to T251 combined with the proximity of N347 to M396, and the high energy conformation of N347 could explain the differences in the emissive output of Akaluc/CouLuc-2-NEt_2 and Akaluc/NapLuc-2-NMe_2. However, further structural studies are warranted to verify this conclusion.

4.5 Conclusions

The use of Rosetta to identify target locations for semi-rational design is a novel technique that has the potential to reduce the number of rounds required for library screening. In the case outlined in this chapter, the two generations of RosettaDesign libraries identified all the necessary mutations to robustly process the luciferin analogues. However, the signal-to-noise ratio in the "hot spot" locations identified by Rosetta was high and leaves room for improvements to the platform. These improvement would likely be found in 1) increasing the constraints of the RosettaDesign step and 2) more exhaustively sampling sequence space through larger design runs.

Using Rosetta to create models to explain experimental results is another novel methodology that proved to be useful in multiple cases. The most notable was the case of identifying hypothesis for the differences in Akaluc's processing of CouLuc-2-NEt_2 and NapLuc-2-NMe_2. One of the keys to success in this arena is a very careful analysis of output models combined with qualitative assessments of said models. However, it must be noted that the hypothesis generated using Rosetta models can only be verified with experimental studies involving structure like crystallography or in some cases, NMR studies. Rosetta models should be used only as a way to generate hypothesis and not to form definite conclusions.

4.6 Acknowledgements

This work was supported by the U.S. National Institutes of Health (R01 GM107630 to J.A.P.) and the Intramural Research Program of the National Institutes of Health (NIH), NCI-CCR. This project has been further supported with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. A.C.L. was supported by the National Science Foundation via the BEST IGERT (DGE-1144901) program. The authors thank the Barbara Karamos Cancer Center for contributing the 4T1, 4T07, and 66Cl4 cancer cell lines, and the NIH NCI-CCR for contributing the JIMT-1 cells. Some experiments were performed at the Laser Spectroscopy labs (LSL) at UCI. The authors thank Dr. Joseph Barchi, NCI-CCR, for NMR assistance and Dr. James Kelley, NCI-CCR, for mass spectrometric analysis. The authors also thank Dr. Jennifer Atwood for performing cell sorting on generated stable cell lines. They acknowledge the support of the Chao Family Comprehensive Cancer Center IFI Flow Cytometry Core Shared Resource, supported by the National Cancer Institute of the National Institutes of Health under award number P30CA062203. Additional thanks to the members of the Prescher lab for helpful discussions, along with members of the Weiss and Martin laboratories for providing equipment and reagents. We also thank Dr. Gary T. Pauly (National Cancer Institute) for assisting with LC/MS and HPLC purification.

CHAPTER 5

FINAL REMARKS

The world of protein design has undergone a tremendous revolution during the period this work was performed. Five years ago, machine learning was a nascent field that had not yet ventured into the protein design space, and the Rosetta software suite dominated. On July 15, 2021, with a massive bang, Google's Deep Mind introduced AlphaFold 2 to the world via *Nature* publishing. That same week, the Baker lab introduced RoseTTAFold via *Science.* To say that this week was a one-two-punch to the tools available to protein designers is a dramatic understatement; the week of July 15, 2021 was a paradigm-shift. Amino acid sequences without structural data were now able to be folded *in silico* with reasonable accuracy. The days of tedious homology modeling to predict protein structure were instantly over. Machine learning was now deeply embedded into the protein design space. Additionally, these tools were made open source, so anyone with a modicum of coding experience could not only use them, but adapt and hack them to address novel challenges.

Since 2021, Rosetta Commons developers, who were previously writing in C++ for the behemoth that is the Rosetta codebase, started shifting away from C++ and toward PyTorch and TensorFlow. In only two years, they developed and open-sourced game-changing machine learning packages like trRosetta, masif (molecular surface interaction fingerprints), ProteinMPNN, RFDiffusion, RFDiffusion-Ligand, and RoseTTAFold-All-Atom. Now, rational protein designers have the ability to create their own scaffold backbones, identify sequences that are likely to adopt the imagined backbone, and calculate the probability of obtaining soluble protein. Amusingly, the Rosetta design suite

of five years ago has been dubbed “Classic Rosetta” in order to differentiate it from the plethora of machine learning-based methodologies associated with the Rosetta Commons.

For the enormous leaps of progress made in the protein design space thanks to machine learning, there are still numerous outstanding challenges. One of these challenges is enzyme design. Recently, Yeh et al. used machine learning methods to design five *de novo* luciferases with function on par to native luciferases (Yeh et al., 2023) without the need for rounds of directed evolution to optimize them. This was an incredible achievement, but the reported success rate for the two rounds of design was 3/7648 and 2/46. This means that between the two rounds of method development, they went from a 0.04% success rate to a 4% success rate. The challenge of designing a well-folded functional enzyme has hardly been solved.

Thus far, machine learning tools have not yet cracked the code to decipher how allosteric interactions affect protein function, and they are unlikely to do so until there are more high-quality site saturation mutagenesis and pairwise mutation datasets available. Unfortunately, these types of data sets are incredibly time consuming and expensive to obtain. Another challenge is that once a large enough data set is obtained, there is no guaranteeing the rules that apply to one family of enzymes will be universal, as reaction mechanisms vary wildly from family to family.

The work presented in this dissertation focused on using “Classic Rosetta” to study the relationship between enzyme substrate specificity and global dynamics. This included a novel approach to targeting residues believed to be important to global dynamics in an effort to alter enzyme function. Moreover, platform development on how to mold an enzyme active site to accommodate novel ligands was undertaken. As

comprehensive as I tried to be, the work presented here is just the beginning, and although I successfully altered the specificity for Fluc by identifying mutations important to remodeling the active site to accommodate larger luciferins, I was unsuccessful in all attempts to alter substrate specificity by targeting allosteric locations distal to the active site in TEM-1. Understanding how allostery and enzyme function are related is an ongoing challenge. I look forward to discovering new ways of approaching the ”dynamics mystery” as the field continues to generate data and develop new hypothesis.

REFERENCES

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., & Gray, J. J. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*, *13*(6), 3031–3048. https://doi.org/10.1021/acs.jctc.7b00125

Ambler, R. P., Coulson, A. F. W., Frère, J. M., Ghuysen, J. M., Joris, B., Forsman, M., Levesque, R. C., Tiraby, G., & Waley, S. G. (1991). A standard numbering scheme for the class A β-lactamases. *Biochemical Journal*, *276*(1), 269–270. https://doi.org/10.1042/bj2760269

Amor, B. R. C., Schaub, M. T., Yaliraki, S. N., & Barahona, M. (2016). Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms12477

Arnold, F. (2019). *Frances H. Arnold -- Nobel Lecture*. Nobel Medai AB 2019. https://www.nobelprize.org/prizes/chemistry/2018/arnold/lecture

Arnold, F. H., Wintrode, P. L., Miyazaki, K., & Gershenson, A. (2001). How enzymes adapt: Lessons from directed evolution. In *Trends in Biochemical Sciences* (Vol. 26, Issue 2, pp. 100–106). https://doi.org/10.1016/S0968-0004(00)01755-2

Atilgan, C., Gerek, Z. N., Ozkan, S. B., & Atilgan, A. R. (2010). Manipulation of conformational change in proteins by single-residue perturbations. *Biophysical Journal*, *99*(3), 933–943. https://doi.org/10.1016/j.bpj.2010.05.020

Baker, D. (2010). An exciting but challenging road ahead for computational enzyme design. *Protein Science*, *19*(10), 1817–1819. https://doi.org/10.1002/pro.481

Bar-Even, A., Milo, R., Noor, E., & Tawfik, D. S. (2015). The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness. *Biochemistry*, *54*, 30. https://doi.org/10.1021/acs.biochem.5b00621

Bar-On, Y. M., & Milo, R. (2019). The global mass and average rate of rubisco. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(10), 4738–4743. https://doi.org/10.1073/pnas.1816654116

Belsare, K. D., Andorfer, M. C., Cardenas, F. S., Chael, J. R., Park, H. J., & Lewis, J. C. (2017). A Simple Combinatorial Codon Mutagenesis Method for Targeted Protein Engineering. *ACS Synthetic Biology*, *6*(3), 416–420. https://doi.org/10.1021/acssynbio.6b00297

Bender, B. J., Cisneros, A., Duran, A. M., Finn, J. A., Fu, D., Lokits, A. D., Mueller, B. K., Sangha, A. K., Sauer, M. F., Sevy, A. M., Sliwoski, G., Sheehan, J. H., DiMaio, F., Meiler, J., & Moretti, R. (2016). Protocols for Molecular Modeling with Rosetta3 and RosettaScripts. *Biochemistry*, *55*(34), 4748–4763. https://doi.org/10.1021/acs.biochem.6b00444

Bishop, C. (2006). *Pattern Recognition and Machine Learning* (1st ed.). Springer-Verlag New York.

Boehr, D. D., Dyson, H. J., & Wright, P. E. (2006). An NMR perspective on enzyme dynamics. *Chemical Reviews*, *106*(8), 3055–3079. https://doi.org/10.1021/cr050312q

Bolon, D. N., & Mayo, S. L. (2001). Enzyme-like proteins by computational design. *Proceedings of the National Academy of Sciences*, *98*(25), 14274–14279. https://doi.org/10.1073/pnas.251555398

Branchini, B. R., Murtiashaw, M. H., Carmody, J. N., Mygatt, E. E., & Southworth, T. L. (2005). Synthesis of an N-acyl sulfamate analog of luciferyl-AMP: A stable and potent inhibitor of firefly luciferase. *Bioorganic & Medicinal Chemistry Letters*, *15*(17), 3860–3864. https://doi.org/10.1016/j.bmcl.2005.05.115

Brandt, C., Braun, S. D., Stein, C., Slickers, P., Ehricht, R., Pletz, M. W., & Makarewicz, O. (2017). In silico serine β-lactamases analysis reveals a huge potential resistome in environmental and pathogenic species. *Scientific Reports*, *7*(January), 1–13. https://doi.org/10.1038/srep43232

Brown, C. A., Hu, L., Sun, Z., Patel, M. P., Singh, S., Porter, J. R., Sankaran, B., Prasad, B. V. V., Bowman, G. R., & Palzkill, T. (2020). Antagonism between substitutions in β-lactamase explains a path not taken in the evolution of bacterial drug resistance. *Journal of Biological Chemistry*, *295*(21), 7376–7390. https://doi.org/10.1074/jbc.RA119.012489

Buchholz, K. (2016). A breakthrough in enzyme technology to fight penicillin resistance—industrial application of penicillin amidase. In *Applied Microbiology and Biotechnology* (Vol. 100, Issue 9, pp. 3825–3839). Springer Verlag. https://doi.org/10.1007/s00253-016-7399-6

Bush, K. (2018). Past and Present Perspectives on β-Lactamases. *Antimicrobial Agents and Chemotherapy*, *62*(10), 1–20. https://doi.org/10.1128/AAC.01076-18

Butler, B. M., Gerek, Z. N., Kumar, S., & Ozkan, S. B. (2015). Conformational dynamics of nonsynonymous variants at protein interfaces reveals disease association. *Proteins: Structure, Function, and Bioinformatics*, *83*(3), 428–435. https://doi.org/10.1002/prot.24748

Campitelli, P., Guo, J., Zhou, H.-X., & Ozkan, S. B. (2018). Hinge-Shift Mechanism Modulates Allosteric Regulations in Human Pin1. *The Journal of Physical Chemistry B*, *122*(21), 5623–5629. https://doi.org/10.1021/acs.jpcb.7b11971

Campitelli, P., Modi, T., Kumar, S., & Ozkan, S. B. (2020). The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annual Review of Biophysics*, *49*(1), 267–288. https://doi.org/10.1146/annurev-biophys-052118-115517

Campitelli, P., Ozkan, S. B., & Swint-Kruse, L. (2020). Asymmetry in Dynamic Allosteric Residue Coupling (DARC) Interactions Captures Evolutionary Landscape. *Biophysical Journal*, *118*(3), 52a. https://doi.org/10.1016/j.bpj.2019.11.464

Chen, C. C. H., Rahil, J., Pratt, R. F., & Herzberg, O. (1993). Structure of a Phosphonate-inhibited β-Lactamase. *Journal of Molecular Biology*, *234*(1), 165–178. https://doi.org/10.1006/jmbi.1993.1571

Chen, K., & Arnold, F. H. (1993). Tuning the activity of an enzyme for unusual environments: Sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. In *Biochemistry Communicated by Peter B. Dervan* (Vol. 90).

Chen, K., Robinson, A. C., Van Dam, M. E., Martinez, P., Economou, C., & Arnold, F. H. (1991). Enzyme Engineering for Nonaqueous Solvents. II. Additive Effects of Mutations on the Stability and Activity of Subtilisin E in Polar Organic Media. In *Biotechnol. Prog* (Vol. 7). https://pubs.acs.org/sharingguidelines

Chica, R. A., Doucet, N., & Pelletier, J. N. (2005). Semi-rational approaches to engineering enzyme activity: Combining the benefits of directed evolution and rational design. In *Current Opinion in Biotechnology* (Vol. 16, Issue 4, pp. 378–384). https://doi.org/10.1016/j.copbio.2005.06.004

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., Biedermannova, L., Sochor, J., & Damborsky, J. (2012). CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Computational Biology*, *8*(10). https://doi.org/10.1371/journal.pcbi.1002708

Clark, D. P., & Pazdernik, N. J. (2008). *Biotechnology: Applying the Genetic Revolution*. Academic Cell.

Conti, E., Franks, N. P., & Brick, P. (1996). Crystal structure of firefly luciferase throws light on a super-family of adenylate-forming enzymes. *Structure*, *4*(3), 287–298. https://doi.org/10.1016/S0969-2126(96)00033-0

Conway, P., Tyka, M. D., DiMaio, F., Konerding, D. E., & Baker, D. (2014a). Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Science*, *23*(1), 47–55. https://doi.org/10.1002/pro.2389

Cortina, G. A., Hays, J. M., & Kasson, P. M. (2018). Conformational Intermediate That Controls KPC-2 Catalysis and Beta-Lactam Drug Resistance. *ACS Catalysis*, *8*(4), 2741–2747. https://doi.org/10.1021/acscatal.7b03832

Cortina, G. A., & Kasson, P. M. (2018). Predicting allostery and microbial drug resistance with molecular simulations. *Current Opinion in Structural Biology*, *52*(1), 80–86. https://doi.org/10.1016/j.sbi.2018.09.001

Coulson, A. (1985). ß-Lactamases: Molecular Studies. *Biotechnology and Genetic Engineering Reviews*, *3*(1), 219–254. https://doi.org/10.1080/02648725.1985.10647814

Dellus-Gur, E., Elias, M., Caselli, E., Prati, F., Salverda, M. L. M., de Visser, J. A. G. M., Fraser, J. S., & Tawfik, D. S. (2015). Negative Epistasis and Evolvability in TEM-1 β-Lactamase—The Thin Line between an Enzyme's Conformational Freedom and Disorder. *Journal of Molecular Biology*, *427*(14), 2396–2409. https://doi.org/10.1016/j.jmb.2015.05.011

Doucet, N., De Wals, P.-Y., & Pelletier, J. N. (2004). Site-saturation Mutagenesis of Tyr-105 Reveals Its Importance in Substrate Stabilization and Discrimination in TEM-1 β-Lactamase. *Journal of Biological Chemistry*, *279*(44), 46295–46303. https://doi.org/10.1074/jbc.M407606200

Doucet, N., Savard, P.-Y., Pelletier, J. N., & Gagné;, S. M. (2007). NMR Investigation of Tyr105 Mutants in TEM-1 β-Lactamase. *Journal of Biological Chemistry*, *282*(29), 21448–21459. https://doi.org/10.1074/jbc.M609777200

Dunbrack, Roland L., Jr. and Karplus, M. (1993). Backbone-dependent rotamer library for proteins. Application to side-chain prediction. In *Journal of Molecular Biology* (Issue 230, pp. 543–574).

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography*, *66*(4), 486–501. https://doi.org/10.1107/S0907444910007493

Fair, R. J., & Tor, Y. (2014). Antibiotics and Bacterial Resistance in the 21st Century. *Perspectives in Medicinal Chemistry*, *6*(6), PMC.S14459. https://doi.org/10.4137/PMC.S14459

Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E. M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J., & Baker, D. (2011).

Rosettascripts: A scripting language interface to the Rosetta Macromolecular modeling suite. *PLoS ONE*, *6*(6). https://doi.org/10.1371/journal.pone.0020161

Forsen, S. (1993). *Nobel Lectures in Chemistry, 1971-1980* (S. Forsen, Ed.). World Scientific Publishing.

Garcia, S. E. (2018). *Nobel Prize in Chemistry Goes to a Woman for the Fifth Time in History*. The New York Times. https://www.nytimes.com/2018/10/3/science/frances-arnold-nobel-prize-chemistry.html

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., B. A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In J. M. Walker (Ed.), *The Proteomic Protocols Handbook* (pp. 571–607). Humana Press.

Gerek, Z. N., Keskin, O., & Ozkan, S. B. (2009). Identification of specificity and promiscuity of PDZ domain interactions through their dynamic behavior. *Proteins: Structure, Function and Bioinformatics*, *77*(4), 796–811. https://doi.org/10.1002/prot.22492

Gerek, Z. N., & Ozkan, S. B. (2011). Change in Allosteric Network Affects Binding Affinities of PDZ Domains: Analysis through Perturbation Response Scanning. *PLoS Comput Biol*, *7*(10), 1002154. https://doi.org/10.1371/journal.pcbi.1002154

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Gobeil, S. M. C., Ebert, M. C. C. J. C., Park, J., Gagné, D., Doucet, N., Berghuis, A. M., Pleiss, J., & Pelletier, J. N. (2019). The Structural Dynamics of Engineered β-Lactamases Vary Broadly on Three Timescales yet Sustain Native Function. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-42866-8

Goldsmith, M., & Tawfik, D. S. (2017). Enzyme engineering: reaching the maximal catalytic efficiency peak. *Current Opinion in Structural Biology*, *47*, 140–150. https://doi.org/10.1016/j.sbi.2017.09.002

Guarnera, E., & Berezovsky, I. N. (2016). Allosteric sites: Remote control in regulation of protein activity. In *Current Opinion in Structural Biology* (Vol. 37, pp. 1–8). https://doi.org/10.1016/j.sbi.2015.10.004

Halgren, T. A. (1996). Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *Journal of Computational Chemistry*, *17*(5–6), 520–552. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<520::AID-JCC2>3.0.CO;2-W

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, *4*(1), 17. https://doi.org/10.1186/1758-2946-4-17

Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., & Stahl, M. T. (2010a). Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling*, *50*(4), 572–584. https://doi.org/10.1021/ci100031x

Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., & Stahl, M. T. (2010b). Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *Journal of Chemical Information and Modeling*, *50*(4), 572–584. https://doi.org/10.1021/ci100031x

Heckmann, C. M., & Paradisi, F. (2020). Looking Back: A Short History of the Discovery of Enzymes and How They Became Powerful Chemical Tools. *ChemCatChem*, *12*(24), 6082–6102. https://doi.org/10.1002/cctc.202001107

Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M., & Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, *450*(7171), 913–916. https://doi.org/10.1038/nature06407

Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, *450*(7172), 964–972. https://doi.org/10.1038/nature06522

Huang, P. S., Boyken, S. E., & Baker, D. (2016). The coming of age of de novo protein design. *Nature*, *537*(7620), 320–327. https://doi.org/10.1038/nature19946

Hughes, S. S. (2011). *Genentech*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226359205.001.0001

Iwano, S., Sugiyama, M., Hama, H., Watakabe, A., Hasegawa, N., Kuchimaru, T., Tanaka, K. Z., Takahashi, M., Ishida, Y., Hata, J., Shimozono, S., Namiki, K., Fukano, T., Kiyama, M., Okano, H., Kizaka-Kondoh, S., McHugh, T. J., Yamamori, T., Hioki, H., … Miyawaki, A. (2018). Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science*, *359*(6378), 935–939. https://doi.org/10.1126/science.aaq1067

Jelsch, C.; Mourey, L.; Masson, J.M.; Samama, J. P. (1993). Crystal structure of Escherichia coli TEM1 beta-lactamase at 1.8 A resolution. *Proteins*, *16*, 364–383.

Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Iii, C. F. B., Hilvert, D., Houk, K. N., Stoddard,

B. L., Baker, D., Barbas, C. F., Hilvert, D., Houk, K. N., Stoddard, B. L., & Baker, D. (2008). De novo computational design of retro-aldol enzymes. *ScienceScience*, *319*(5868), 1387–1391. https://doi.org/10.1126/science.1152692
Kabsch, W. (2010a). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D Biological Crystallography*, *66*(2), 133–144. https://doi.org/10.1107/S0907444909047374

Kabsch, W. (2010b). XDS. *Acta Crystallographica Section D Biological Crystallography*, *66*(2), 125–132. https://doi.org/10.1107/S0907444909047337

Kamerlin, S. C. L., & Warshel, A. (2010). At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis. *Proteins: Structure, Function and Bioinformatics*, *78*(6), 1339–1375. https://doi.org/10.1002/prot.22654

Kaskova, Z. M., Tsarkova, A. S., & Yampolsky, I. V. (2016). 1001 lights: Luciferins, luciferases, their mechanisms of action and applications in chemical analysis, biology and medicine. *Chemical Society Reviews*, *45*(21), 6048–6077. https://doi.org/10.1039/c6cs00296j

Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H., & Meiler, J. (2010). Practically useful: What the R osetta protein modeling suite can do for you. *Biochemistry*, *49*(14), 2987–2998. https://doi.org/10.1021/bi902153g

Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., Popović, Z., Baker, D., & Players, F. (2011). Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences*, *108*(47), 18949–18953. https://doi.org/10.1073/pnas.1115898108

Kim, H., Zou, T., Modi, C., Dörner, K., Grunkemeyer, T. J., Chen, L., Fromme, R., Matz, M. V., Ozkan, S. B., & Wachter, R. M. (2015). A Hinge Migration Mechanism Unlocks the Evolution of Green-to-Red Photoconversion in GFP-like Proteins. *Structure*, *23*(1), 34–43. https://doi.org/10.1016/j.str.2014.11.011

Kiss, G., Röthlisberger, D., Baker, D., & Houk, K. N. (2010). Evaluation and ranking of enzyme designs. *Protein Science*, *19*(9), 1760–1773. https://doi.org/10.1002/pro.462

Korendovych, Ivan V., DeGrado, W. F. (2014). Catalytic efficiency of designed catalytic proteins. *Current Opinion in Structural Biology*, *23*(0), 113–121. https://doi.org/10.1016/j.sbi.2014.06.006.

Kortemme, T., Morozov, A. V., & Baker, D. (2003). An Orientation-dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein–Protein Complexes. *Journal of Molecular Biology*, *326*(4), 1239–1259. https://doi.org/10.1016/S0022-2836(03)00021-4

Koshland, D. E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences*, *44*(2), 98–104. https://doi.org/10.1073/pnas.44.2.98
Kovalevskiy, O., Nicholls, R. A., Long, F., Carlon, A., & Murshudov, G. N. (2018). Overview of refinement procedures within *REFMAC* 5: utilizing data from different sources. *Acta Crystallographica Section D Structural Biology*, *74*(3), 215–227. https://doi.org/10.1107/S2059798318000979

Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., & Baker, D. (2003). Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science*, *302*(5649), 1364–1368. https://doi.org/10.1126/science.1089427

Kumar, A., Butler, B. M., Kumar, S., & Ozkan, S. B. (2015). Integration of structural dynamics and molecular evolution via protein interaction networks: a new era in genomic medicine. *Current Opinion in Structural Biology*, *35*(3), 135–142. https://doi.org/10.1016/j.sbi.2015.11.002

Kumar, A., Glembo, T. J., & Ozkan, S. B. (2015). The Role of Conformational Dynamics and Allostery in the Disease Development of Human Ferritin. *Biophysical Journal*, *109*(6), 1273–1281. https://doi.org/10.1016/j.bpj.2015.06.060

Larrimore, K. E., Kazan, I. C., Kannan, L., Kendle, R. P., Jamal, T., Barcus, M., Bolia, A., Brimijoin, S., Zhan, C. G., Ozkan, S. B., & Mor, T. S. (2017). Plant-expressed cocaine hydrolase variants of butyrylcholinesterase exhibit altered allosteric effects of cholinesterase activity and increased inhibitor sensitivity. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-10571-z

Lazaridis, T. (2003). Effective energy function for proteins in lipid membranes. *Proteins: Structure, Function and Genetics*, *52*(2), 176–192. https://doi.org/10.1002/prot.10410

Leaver-Fay, A., Jacak, R., Stranges, P. B., & Kuhlman, B. (2011). A Generic Program for Multistate Protein Design. *PLoS ONE*, *6*(7), e20937. https://doi.org/10.1371/journal.pone.0020937

Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y. E. A., Fleishman, S. J., Corn, J. E., Kim, D. E., … Bradley, P. (2011). Rosetta3: An object-oriented software suite for the simulation and design of macromolecules. In *Methods in Enzymology* (Vol. 487, Issue C, pp. 545–574). Academic Press Inc. https://doi.org/10.1016/B978-0-12-381270-4.00019-6

Leferink, N. G. H., Antonyuk, S. V, Houwman, J. A., Scrutton, N. S., Eady, R. R., & Hasnain, S. S. (2014). Impact of residues remote from the catalytic centre on enzyme catalysis of copper nitrite reductase. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms5395

Lichtenthaler, F. W. (1995). 100 Years “Schlüssel-Schloss-Prinzip”: What Made Emil Fischer Use this Analogy? *Angewandte Chemie International Edition in English*, *33*(23–24), 2364–2374. https://doi.org/10.1002/anie.199423641

Lindskog, S., & Colemant, J. E. (1973). *The Catalytic Mechanism of Carbonic Anhydrase (metalloenzymes/enzyme mechanism/hydration of C02)* (Vol. 70, Issue 9).

Love, A. C., Caldwell, D. R., Kolbaba-Kartchner, B., Townsend, K. M., Halbers, L. P., Yao, Z., Brennan, C. K., Ivanic, J., Hadjian, T., Mills, J. H., Schnermann, M. J., & Prescher, J. A. (2023). Red-Shifted Coumarin Luciferins for Improved Bioluminescence Imaging. *Journal of the American Chemical Society*, *145*(6), 3335–3345. https://doi.org/10.1021/jacs.2c07220

Ma, B., Kumar, S., Tsai, C.-J., & Nussinov, R. (1999). Folding funnels and binding mechanisms. In *Protein Engineering vol* (Vol. 12, Issue 9).

Ma, B., Tsai, C.-J., Haliloğlu, T., & Nussinov, R. (2011). Dynamic Allostery: Linkers Are Not Merely Flexible. *Structure*, *19*(7), 907–917. https://doi.org/10.1016/j.str.2011.06.002

Maguire, J. B., Haddox, H. K., Strickland, D., Halabiya, S. F., Coventry, B., Griffin, J. R., Pulavarti, S. V. S. R. K., Cummins, M., Thieker, D. F., Klavins, E., Szyperski, T., DiMaio, F., Baker, D., & Kuhlman, B. (2021). Perturbing the energy landscape for improved packing during computational protein design. *Proteins: Structure, Function, and Bioinformatics*, *89*(4), 436–449. https://doi.org/10.1002/prot.26030

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, *11*(8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255

Mak, W. S., & Siegel, J. B. (2014). Computational enzyme design: Transitioning from catalytic proteins to enzymes. In *Current Opinion in Structural Biology* (Vol. 27, Issue 1, pp. 87–94). https://doi.org/10.1016/j.sbi.2014.05.010

Markin, C. J., Mokhtari, D. A., Sunden, F., Appel, M. J., Akiva, E., Longwell, S. A., Sabatti, C., Herschlag, D., & Fordyce, P. M. (2021). Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics. *Science*, *373*(6553). https://doi.org/10.1126/science.abf8761

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(4), 658–674. https://doi.org/10.1107/S0021889807021206

Mills, J. H., Khare, S. D., Bolduc, J. M., Forouhar, F., Mulligan, V. K., Lew, S., Seetharaman, J., Tong, L., Stoddard, B. L., & Baker, D. (2013). Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. *Journal of the American Chemical Society*, *135*(36), 13393–13399. https://doi.org/10.1021/ja403503m

Mills, J. H., Sheffler, W., Ener, M. E., Almhjell, P. J., Oberdorfer, G., Pereira, José Henrique Parmeggiani, F., Sankaran, B., Zwart, P. H., & Baker, D. (2016). Computational design of a homotrimeric metalloprotein with a trisbipyridyl core. *Proceedings of the National Academy of Sciences*, *113*(52), 15012–15017. https://doi.org/10.1073/pnas.1600188113

Modi, T., & Banu Ozkan, S. (2018). Mutations utilize dynamic allostery to confer resistance in TEM-1 β-lactamase. *International Journal of Molecular Sciences*, *19*(12). https://doi.org/10.3390/ijms19123808

Modi, T., Campitelli, P., Kazan, I. C., & Ozkan, S. B. (2021). Protein folding stability and binding interactions through the lens of evolution: a dynamical perspective. *Current Opinion in Structural Biology*, *66*, 207–215. https://doi.org/10.1016/j.sbi.2020.11.007

Modi, T., Huihui, J., Ghosh, K., & Ozkan, S. B. (2018). Ancient thioredoxins evolved to modern-day stability–function requirement by altering native state ensemble. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1749). https://doi.org/10.1098/rstb.2017.0184

Monod, J., & Jacob, F. (1978). General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. In *Selected Papers in Molecular Biology by Jacques Monod* (pp. 491–503). Elsevier. https://doi.org/10.1016/B978-0-12-460482-7.50044-0

Morley, K. L., & Kazlauskas, R. J. (2005). Improving enzyme properties: When are closer mutations better? *Trends in Biotechnology*, *23*(5), 231–237. https://doi.org/10.1016/j.tibtech.2005.03.005

Mullis, K., Ferre, F., & Gibbs, R. A. (1994). *The polymerase chain reaction*. Birkhauser.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 355–367. https://doi.org/10.1107/S0907444911001314

Murshudov, G. N., Vagin, A. A., & Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D Biological Crystallography*, *53*(3), 240–255. https://doi.org/10.1107/S0907444996012255

Nakatsu, T., Ichiyama, S., Hiratake, J., Saldanha, A., Kobashi, N., Sakata, K., & Kato, H. (2006). Structural basis for the spectral difference in luciferase bioluminescence. *Nature*, *440*(7082), 372–376. https://doi.org/10.1038/nature04542
NCBI BioProject Database. (n.d.). https://www.ncbi.nlm.nih.gov/bioproject/

Nevin Gerek, Z., Kumar, S., & Banu Ozkan, S. (2013). Structural dynamics flexibility informs function and evolution at a proteome scale. *Evolutionary Applications*, *6*(3), 423–433. https://doi.org/10.1111/eva.12052

Nicholls, R. A., Tykac, M., Kovalevskiy, O., & Murshudov, G. N. (2018). Current approaches for the fitting and refinement of atomic models into cryo-EM maps using *CCP-EM*. *Acta Crystallographica Section D Structural Biology*, *74*(6), 492–505. https://doi.org/10.1107/S2059798318007313

Nivón, L. G., Moretti, R., & Baker, D. (2013). A Pareto-Optimal Refinement Method for Protein Design Scaffolds. *PLoS ONE*, *8*(4), e59004. https://doi.org/10.1371/journal.pone.0059004

Nobel Foundation. (1964). *Nobel Lectures, Chemistry 1942-1962*. Elsevier Publishing Company.
Ojeda-May, P., Mushtaq, A. U., Rogne, P., Verma, A., Ovchinnikov, V., Grundström, C., Dulko-Smith, B., Sauer, U. H., Wolf-Watz, M., & Nam, K. (2021). Dynamic Connection between Enzymatic Catalysis and Collective Protein Motions. *Biochemistry*, *60*(28), 2246–2258. https://doi.org/10.1021/acs.biochem.1c00221

Ollikainen, N., de Jong, R. M., & Kortemme, T. (2015). Coupling Protein Side-Chain and Backbone Flexibility Improves the Re-design of Protein-Ligand Specificity. *PLoS Computational Biology*, *11*(9), 1–22. https://doi.org/10.1371/journal.pcbi.1004335

Orencia, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P. C., & Stevens, R. C. (2001). Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nature Structural Biology*, *8*(3), 238–242. https://doi.org/10.1038/84981

Paul, F., & Weikl, T. R. (2016). How to Distinguish Conformational Selection and Induced Fit Based on Chemical Relaxation Rates. *PLoS Computational Biology*, *12*(9). https://doi.org/10.1371/journal.pcbi.1005067

Pauling, L. (1946). Molecular Architecture and Biological Reactions. *Chemical & Engineering News Archive*, *24*(10), 1375–1377. https://doi.org/10.1021/cen-v024n010.p1375

Privett, H. K., Kiss, G., Lee, T. M., Blomberg, R., Chica, R. A., Thomas, L. M., Hilvert, D., Houk, K. N., & Mayo, S. L. (2012). Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences*, *109*(10), 3790–3795. https://doi.org/10.1073/pnas.1118082108

Radzika, A., & Wolfenden, R. (1995). A Proficient Enzyme. *Science*, *267*(1968), 90–93.

Ramanathan, A., & Agarwal, P. K. (2011). Evolutionarily Conserved Linkage between Enzyme Fold, Flexibility, and Catalysis. *PLoS Biology*, *9*(11), e1001193. https://doi.org/10.1371/journal.pbio.1001193

Ramanathan, A., Savol, A. J., Langmead, C. J., Agarwal, P. K., & Chennubhotla, C. S. (2011). Discovering Conformational Sub-States Relevant to Protein Function. *PLoS ONE*, *6*(1), e15827. https://doi.org/10.1371/journal.pone.0015827

Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A., & Skiff, W. M. (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, *114*(25), 10024–10035. https://doi.org/10.1021/ja00051a040

Renata, H., Wang, Z. J., & Arnold, F. H. (2015). Expanding the enzyme universe: Accessing non-natural reactions by mechanism-guided directed evolution. In *Angewandte Chemie - International Edition* (Vol. 54, Issue 11). https://doi.org/10.1002/anie.201409470

Rice, B. W., Cable, M. D., & Nelson, M. B. (2001). In vivo imaging of light-emitting probes. *Journal of Biomedical Optics*, *6*(4), 432. https://doi.org/10.1117/1.1413210

Richter, F., Blomberg, R., Khare, S. D., Kiss, G., Kuzin, A. P., Smith, A. J. T., Gallaher, J., Pianowski, Z., Helgeson, R. C., Grjasnow, A., Xiao, R., Seetharaman, J., Su, M., Vorobiev, S., Lew, S., Forouhar, F., Kornhaber, G. J., Hunt, J. F., Montelione, G. T., … Baker, D. (2012). Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *Journal of the American Chemical Society*, *134*(39), 16197–16206. https://doi.org/10.1021/ja3037367

Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S., & Baker, D. (2011). De novo enzyme design using Rosetta3. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0019230

Ringer, A. L., Senenko, A., & Sherrill, C. D. (2007). Models of S/π interactions in protein structures: Comparison of the H2 $_2$ S–benzene complex with PDB data. *Protein Science*, *16*(10), 2216–2223. https://doi.org/10.1110/ps.073002307

Risso, V. A., Gavira, J. A., Gaucher, E. A., & Sanchez-Ruiz, J. M. (2014). Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins. *Proteins: Structure, Function and Bioinformatics*, *82*(6), 887–896. https://doi.org/10.1002/prot.24575

Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A., & Sanchez-Ruiz, J. M. (2013). Hyperstability and substrate promiscuity in laboratory resurrections of

precambrian β-lactamases. *Journal of the American Chemical Society*, *135*(8), 2899–2902. https://doi.org/10.1021/ja311630a

Rohl, C. A., Strauss, C. E. M. M., Misura, K. M. S. S., & Baker, D. (2004). Protein Structure Prediction Using Rosetta. *Methods in Enzymology*, *383*(2003), 66–93. https://doi.org/10.1016/S0076-6879(04)83004-0

Röthlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., Dechancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., & Baker, D. (2008). *Kemp elimination catalysts by computational enzyme design*. *453*. https://doi.org/10.1038/nature06879

Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., & Walker, R. C. (2013). 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. https://doi.org/10.1021/ct400314y

Salverda, M. L. M., De Visser, J. A. G. M., & Barlow, M. (2010). Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiology Reviews*, *34*(6), 1015–1036. https://doi.org/10.1111/j.1574-6976.2010.00222.x

Schneider, S. H., Kozuch, J., & Boxer, S. G. (2021). *The Interplay of Electrostatics and Chemical Positioning in the Evolution of Antibiotic Resistance in TEM β -Lactamases*. https://doi.org/10.1101/2021.05.27.446023

Schnell, J. R., Dyson, H. J., & Wright, P. E. (2004). Structure, dynamics, and catalytic function of dihydrofolate reductase. In *Annual Review of Biophysics and Biomolecular Structure* (Vol. 33, pp. 119–140). https://doi.org/10.1146/annurev.biophys.33.110502.133613

Schramm, V. L. (2011). Enzymatic Transition States, Transition-State Analogs, Dynamics, Thermodynamics, and Lifetimes. *Annual Review of Biochemistry*, *80*(1), 703–732. https://doi.org/10.1146/annurev-biochem-061809-100742

Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St.Clair, J. L., Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E., & Baker, D. (2010). Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science*, *329*(5989), 309–313. https://doi.org/10.1126/science.1188934

Sielecki, A., Betzel, C., James, M. N. G., Adachi, H., Strynadka, N. C. J., Jensen, S. E., Sutoh, K., & Johns, K. (2003). Molecular structure of the acyl-enzyme intermediate in β-lactam hydrolysis at 1.7 Å resolution. *Nature*, *359*(6397), 700–705. https://doi.org/10.1038/359700a0

Singh, M. K., & Dominy, B. N. (2012). The Evolution of Cefotaximase Activity in the TEM β-Lactamase. *Journal of Molecular Biology*, *415*(1), 205–220. https://doi.org/10.1016/j.jmb.2011.10.041

Singh, P., Francis, K., & Kohen, A. (2015). Network of remote and local protein dynamics in dihydrofolate reductase catalysis. *ACS Catalysis*, *5*(5), 3067–3073. https://doi.org/10.1021/acscatal.5b00331

Singh, R., Kumar, M., Mittal, A., & Mehta, P. K. (2016). Microbial enzymes: industrial progress in 21st century. *3 Biotech*, *6*(2), 1–15. https://doi.org/10.1007/s13205-016-0485-8

Smith, A. J. T., Müller, R., Toscano, M. D., Kast, P., Hellinga, H. W., Hilvert, D., & Houk, K. N. (2008). 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. https://doi.org/10.1021/ja803213p

Song, Z., Zhang, Q., Wu, W., Pu, Z., & Yu, H. (2023). Rational design of enzyme activity and enantioselectivity. In *Frontiers in Bioengineering and Biotechnology* (Vol. 11). Frontiers Media S.A. https://doi.org/10.3389/fbioe.2023.1129149

Stiffler, M. A., Hekstra, D. R., & Ranganathan, R. (2015). Evolvability as a Function of Purifying Selection in TEM-1 β-Lactamase. *Cell*, *160*(5), 882–892. https://doi.org/10.1016/j.cell.2015.01.035

Sumner, J. B., & Somers, G. F. (2014). *Chemistry and Methods of Enzymes* (3rd ed.). Academic Press.

Sundlov, J. A., Fontaine, D. M., Southworth, T. L., Branchini, B. R., & Gulick, A. M. (2012). Crystal Structure of Firefly Luciferase in a Second Catalytic Conformation Supports a Domain Alternation Mechanism. *Biochemistry*, *51*(33), 6493–6495. https://doi.org/10.1021/bi300934s

Tantillo, D. J., Chen, J., & Houk, K. N. (1998). Theozymes and compuzymes: theoretical models for biological catalysis. *Current Opinion in Chemical Biology*, *2*, 743–750. http://biomednet.com/elecref/1367593100200743

The PyMOL Molecular Graphics System, Version 4.3 Schrödinger, LLC. (n.d.).

Tinberg, C. E., & Khare, S. D. (2017). Computational design of ligand binding proteins. *Methods in Molecular Biology*, *1529*(44), 363–373. https://doi.org/10.1007/978-1-4939-6637-0_19

Tiwari, M. K., Singh, R., Singh, R. K., Kim, I. W., & Lee, J. K. (2012). Computational approaches for rational design of proteins with novel functionalities. *Computational and Structural Biotechnology Journal*, *2*(3), e201204002. https://doi.org/10.5936/csbj.201209002

Tugarinov, V., Ceccon, A., & Clore, G. M. (2021). Probing Side-Chain Dynamics in Proteins by NMR Relaxation of Isolated ^{13} C Magnetization Modes in ^{13} CH $_3$ Methyl Groups. *The Journal of Physical Chemistry B*, *125*(13), 3343–3352. https://doi.org/10.1021/acs.jpcb.1c00989

Tyka, M. D., Keedy, D. a, André, I., DiMaio, F., Song, Y., Richardson, D. C., Richardson, J. S., & Baker, D. (2011). Alternate States of Proteins Revealed by Detailed Energy Landscape Mapping. *Journal of Molecular Biology*, *405*(2), 607–618. https://doi.org/10.1016/j.jmb.2010.11.008

Vaissier Welborn, V., & Head-Gordon, T. (2019). Computational Design of Synthetic Enzymes [Review-article]. *Chemical Reviews*, *119*(11), 6613–6630. https://doi.org/10.1021/acs.chemrev.8b00399

Wang, X., Minasov, G., & Shoichet, B. K. (2002). Evolution of an Antibiotic Resistance Enzyme Constrained by Stability and Activity Trade-offs. *Journal of Molecular Biology*, *320*(1), 85–95. https://doi.org/10.1016/S0022-2836(02)00400-X

Warshel, A., & Bora, R. P. (2016). Perspective: Defining and quantifying the role of dynamics in enzyme catalysis. *Journal of Chemical Physics*, *144*(18), 180901. https://doi.org/10.1063/1.4947037

Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Liu, H., & Olsson, M. H. M. (2006). Electrostatic basis for enzyme catalysis. *Chemical Reviews*, *106*(8), 3210–3235. https://doi.org/10.1021/cr0503106

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Hanikel, N., Pellock, S. J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S. V., … Baker, D. (2023). De novo design of protein structure and function with RFdiffusion. *Nature*. https://doi.org/10.1038/s41586-023-06415-8

Wiegand, I., Hilpert, K., & Hancock, R. E. W. (2008). *Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances*. https://doi.org/10.1038/nprot.2007.521

Williams, S. J., Gewing-Mullins, J. A., Lieberman, W. K., Kolbaba-Kartchner, B., Iqbal, R., Burgess, H. M., Colee, C. M., Ornelas, M. Y., Reid-McLaughlin, E. S., Mills, J. H., Prescher, J. A., & Leconte, A. M. (2023). Biochemical Analysis Leads to Improved

Orthogonal Bioluminescent Tools. *ChemBioChem*, *24*(6). https://doi.org/10.1002/cbic.202200726

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography*, *67*(4), 235–242. https://doi.org/10.1107/S0907444910045749

Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z., & Kern, D. (2004a). Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nature Structural & Molecular Biology*, *11*(10), 945–949. https://doi.org/10.1038/nsmb821

Yao, Z., Zhang, B. S., & Prescher, J. A. (2018). Advances in bioluminescence imaging: new probes from old recipes. *Current Opinion in Chemical Biology*, *45*, 148–156. https://doi.org/10.1016/j.cbpa.2018.05.009

Yeh, A. H. W., Norn, C., Kipnis, Y., Tischer, D., Pellock, S. J., Evans, D., Ma, P., Lee, G. R., Zhang, J. Z., Anishchenko, I., Coventry, B., Cao, L., Dauparas, J., Halabiya, S., DeWitt, M., Carter, L., Houk, K. N., & Baker, D. (2023). De novo design of luciferases using deep learning. *Nature*, *614*(7949), 774–780. https://doi.org/10.1038/s41586-023-05696-3

Zanghellini, A., Jiang, L., Wollacott, A. M., Cheng, G., Meiler, J., Althoff, E. A., Röthlisberger, D., & Baker, D. (2006). New algorithms and an in silico benchmark for computational enzyme design. *Protein Science*, *15*(12), 2785–2794. https://doi.org/10.1110/ps.062353106

Zeymer, C., Werbeck, N. D., Zimmermann, S., Reinstein, J., & Hansen, D. F. (2016). Characterizing Active Site Conformational Heterogeneity along the Trajectory of an Enzymatic Phosphoryl Transfer Reaction. *Angewandte Chemie*, *128*(38), 11705–11709. https://doi.org/10.1002/ange.201606238

Zhang, Y., Doruker, P., Kaynak, B., Zhang, S., Krieger, J., Li, H., & Bahar, I. (2020). Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior. *Current Opinion in Structural Biology*, *62*, 14–21. https://doi.org/10.1016/j.sbi.2019.11.002

Zhu, X., & Lai, L. (2009). A novel method for enzyme design. *Journal of Computational Chemistry*, *30*(2), 256–267. https://doi.org/10.1002/jcc.21050

Zou, T., Risso, V. A., Gavira, J. A., Sanchez-Ruiz, J. M., & Ozkan, S. B. (2015). Evolution of Conformational Dynamics Determines the Conversion of a Promiscuous

Generalist into a Specialist Enzyme. *Molecular Biology and Evolution*, *32*(1), 132–143. https://doi.org/10.1093/molbev/msu281

APPENDIX A

ADDITIONAL PUBLICATIONS

In addition to the publications mentioned specifically in the chapters, I also contributed to the following projects. First, I did the molecular cloning, protein expression and purification of the streptavidin mutants in the following publication. Gleason, P.R., Kolbaba-Kartchner, B., Henderson, J. N., Stahl, E. P., Simmons, C. R., & Mills, J. H. (2021). Structural Origins of Altered Spectroscopic Properties upon Ligand Binding in Proteins Containing a Fluorescent Noncanonical Amino Acid. *Biochemistry (Easton)*, *60*(34), 2577–2585.

Second, I used the RosettaDesign platform to model experimental data in the following publication. Williams, S., Gewing-Mullins, J. A., Lieberman, W. K., Kolbaba-Kartchner, B., Iqbal, R., Burgess, H. M., Colee, C. M., Ornelas, M. Y., Reid-McLaughlin, E. S., Mills, J. H., Prescher, J. A., & Leconte, A. M. (2023). Biochemical Analysis Leads to Improved Orthogonal Bioluminescent Tools. *Chembiochem : a European Journal of Chemical Biology*, *24*(6), e202200726–n/a.

Third, I used the RosettaDesign platform to identify target locations for CCM luciferase library design in the following publication. Yao, Z., Caldwell, D. R., Love, A. C., Kolbaba-Kartchner, B., Mills, J. H., Schnermann, M. J., & Prescher, J. A. (2021). Coumarin luciferins and mutant luciferases for robust multi-component bioluminescence imaging. *Chemical Science (Cambridge)*, *12*(35), 11684–11691.

Fourth, I used the RosettaDesign platform to model experimental data in the following publication which is under review at *The Journal of Organic Chemistry* at the time of this writing. Caldwell, Donald; Townsend, Katherine; Kolbaba-Kartchner, Bethany; Hadjian, Tanya; Ivanic, Joseph; Mills, Jeremy; Prescher, Jennifer; Schnermann, Martin. (2023). Expedient Synthesis and Characterization of π-Extended Luciferins. *The*

Journal of Organic Chemistry.

APPENDIX B

GENERAL ROSETTA SCRIPTS FROM CHAPTER 2

All calculations were carried out using Rosetta version: 442bff4fb7bf2ccb44655e8d15276c9bccfbbd0.

The following command line was used to minimize the total energy of the 1btl crystal structure from the Protein Data Bank using the Rosetta relax protocol:

```
/Rosetta/main/source/bin/relax.default.linuxgccrelease -s pdb_file -
nstruct 1 -relax:default_repeats 5 -
relax:constrain_relax_to_start_coords -
relax:coord_constrain_sidechains -relax:ramp_constraints false -ex1
-ex2 -use_input_sc -flip_HNQ -ignore_unrecognized_res -
relax:coord_cst_stdev 0.5
```

The DesignAround protocol was initiated with the following command line:

```
/Rosetta/main/source/bin/rosetta_scripts.linuxgccrelease -
out:nstruct 25 -jd2:ntrials 50 -parser:protocol /design.xml -
packing:resfile [resfile] -database /Rosetta/main/database -
out::overwrite -s pdb_file -run:preserve_header -output_virtual true
-use_input_sc -no_his_his_pairE -score::hbond_params correct_params
-lj_hbond_hdis 1.75 -lj_hbond_OH_donor_dis 2.6 -linmem_ig 10 -
nblist_autoupdate true -in:ignore_unrecognized_res -out::overwrite
And the contents of design.xml was: ”/> ” resnums=”” repack_shell=””
allow_design=”1” resnums_allow_design=”0”/>
```

The content of the resfile was:

```
ALLAA
EX 1
EX 2
USE_INPUT_SC
Start
2 A PIKAA P
19 A PIKAA V #rigid resi
20 A PIKAA G #rigid resi
37 A PIKAA P
42 A PIKAA P
45 A PIKAA S #Active site
48 A PIKAA K #Active site
51 A PIKAA L #rigid resi
82 A PIKAA P
97 A PIKAA L #rigid resi
105 A PIKAA S #Active site
107 A PIKAA N #Active site
120 A PIKAA P
141 A PIKAA E #Active site
142 A PIKAA P #This proline is important for folding stability
149 A PIKAA P
158 A PIKAA P
209 A PIKAA K #Active site
218 A PIKAA R #Active site
```

```
230 A PIKAA P
235 A PIKAA V #rigid resi
```

Sequences of Designed Proteins in FASTA format

>Native β-lactamase signal peptide MSIQHFRVALIPFFAAFCLPVFA

>Rdg262a
HPETLVKVKDAEDQLGARVGFQLTDLNSGKILEYFRAEERFPMMSTFKVLLCGA
VLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNT
AANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTQPKAMA
QTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIACKSGAGERG
SRGIIAALGPDGKPSRIVVIFTTGSQATMDERNRQIAEIGASLIKHW

>Rdg262b
HPETLVKVKDAEDQLGARVGFILLDLNSGKILESFRAEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTTPRAMAT
TLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERGS
RGQIAALGPDGKPSRIVVIMTTGSQATMDERNRQIAEIGASLIKHW

>Rdg44a
HPETLVKVKDAVDQLGAPVGMIELDLNSGKILESYNPEERFPMMSTFKVLLCGV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT
TLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERGS
RGIIAALGPDGKPSRIVVIMMTGSQATMDERNRAIAEIGASLIKHW

>Rdg44b
HPETLVKVKKAVDDLGAPVGFIELDLNSGKILESYKPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT
TLRKLLTGELLTLASRQQLIDWMEADVAGPLLRSALPAGWFIADKSGAGERGSR
GIIAALGPDGKPSRIVVTMTSGSQATMDERNRAIAEIGASLIKHW

>Rdg44c
HPETLVVVKQAEDKLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT
TLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERGSI
GIIAALGPDGKPSRIVVIYATGSQATMDELNRAIAEIGASLIKHW

>Flx226a
HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT

TLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERGS
RGIIAALGPNGKPSRIVVIYTTGSQATMDERNRQIAEIGASLFKHW

>Flx226b
HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT
TLRKLLTGELLTLASRQQLIDWMEADVAGPLLRSALPPGWFIADKSGAGERGSR
GIIAAFGPNGVPTRIVVIYTTGSQATMDERNRQIAEIGASLFKHY

>Flx226c
HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMATT
LRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPPGWFIADKSGAGERGSR
GIIAALGPNGVPSRIVVIYTTGSQATMDERNRQIAEIGASLFKHW

>Flx256
HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAV
LSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTA
ANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMAT
TLRKLLTGELLTLASRQQLIDWMAADKVAGPLLRSALPPGWFIADKSGAGERGS
RGIIASLGPNGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW

>Flx55
HPETLVKVKDAEDQLGARVGYILLDADSGKILEAFRPEERFPMMSTFKVLLCGA
VLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNT
AANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPRAMA
ETLRKLLLGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERG
SRGIIAMLGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW

APPENDIX B
GENERAL ROSETTA SCRIPTS FROM CHAPTER 5

All calculations were carried out using the Rosetta software suite (version 2021.29+master.d8f55669792) with the ref15 score function (Alford et al., 2017a; Leaver-Fay, Tyka, et al., 2011).

Preparing the scaffolds

A high-resolution (2.62 Å) crystal structure of *Photinus pyralis* luciferase (PDB ID: 4g36) was processed to remove waters and chain B (Sundlov et al., 2012). The structure was subjected to rounds of energy minimization using Rosetta FastRelax with the native co-crystallized ligand, 5′-O-[N-(dehydroluciferyl)-sulfamoyl] adenosine (SLU) because it was noted that removal of the ligand prior to "relaxing" the protein caused motion in a flexible loop in the active site, precluding productive ligand binding in subsequent modeling efforts (Khatib et al., 2011; Maguire et al., 2021; Nivón et al., 2013).The params file used to input the geometric coordinates of each ligand atom, along with the bonding patterns of the atoms for SLU can be found in Appendix D.

The script used for the FastRelax application is as follows.

```
<Pathto>/Rosetta/main/source/bin/relax.default.linuxgccrelease -s
<input_file> -nstruct 5 -relax:default_repeats 5 -
relax:constrain_relax_to_start_coords -ex1 -ex2 -use_input_sc -flip_HNQ
-ignore_unrecognized_res -relax:coord_cst_stdev 0.3 -extra_res_fa
SLU.params
```

The lowest energy model as calculated by the Rosetta score function with a root-mean-square-deviation (RMSD) of < 2.0 Å from the input crystal structure was chosen and the SLU molecule was removed for all downstream applications.

Preparing the ligands

To predict the binding modes of the luciferin analogues, two models of each of the CouLuc-3 and CouLuc-2 ligands (corresponding to the Z and E isomers), one model of the NapLuc analogues, and one model of the AkaLumine analogues were built in Avogadro: an open-source molecular builder and visualization tool, version 1.2.0. http://avogadro.cc/ using the nonhydrolyzable adenyl moiety of SLU as the base. (Hanwell et al., 2012) Each molecule was subjected to an energy minimization in Avogadro using the UFF force field (Rappe et al., 1992) Ligand conformers of both copies of each ligand were generated using OMEGA 4.1.2.0 (OpenEye Scientific Software, Santa Fe, NM) (Hawkins et al., 2010a). During conformer generation, torsion angle sampling was limited to the extended pi system of the novel luciferins by prohibiting movement of the torsions on the adenyl moiety. An example command line for generating conformer libraries of each ligand is as follows. The SMILES string varied slightly with each ligand.

```
oeomega pose -in <input.mol2> -out output.mol2 -ewindow 12 -fixsmarts
"C(=O)NS(=O)(=O)OCC1OC(C(C1O)O)[N+1]=2C=3C(=NC=2)[C+1](N=CN=3)N" -
fromCT false -enumNitrogen false -enumRing false -searchff  mmff_Trunc
-strictatomtyping false
```

The Omega output was a .mol2 file containing the conformers, which was subsequently converted into a .params file with the identities and geometries of each atom specified; a .pdb file containing the library of conformers using an internal Rosetta script was also generated. The params files used to model the luciferin analogues can be found in Appendix D.

General RosettaDesign Methods

The prepared Fluc (PDB ID: 4g36) structure was used as an input to the RosettaMatch protocol, which was used to dock the novel luciferins into the protein

structure based on user-defined constraints (Richter et al., 2011)The following command line was used to call the RosettaMatch application.

```
<Path to>/Rosetta/main/source/bin/match.linuxgccrelease -s <input_file>
-extra_res_fa <params file> -match:geometric_constraint_file
<constraint file> -match:lig_name <ligand 3 letter code>
-match:scaffold_active_site_residues_for_geomcsts <pos_file> -ex1 -ex2
-ex2aro -ex1aro -extrachi_cutoff 0 -use_input_sc true -database
<path_to_Rosetta>/database/ -match:filter_colliding_upstream_residues -
match:filter_upstream_downstream_collisions -
match:upstream_residue_collision_tolerance 0.95 -
match:updown_collision_tolerance 0.3 -match::bump_tolerance 0.3 -
match_grouper SameSequenceAndDSPositionGrouper -
match:grouper_downstream_rmsd 0.5 -match:euclid_bin_size 0.5 -
match:euler_bin_size 5.0 -output_format PDB -exclude_patches
N_acetylated -consolidate_matches 1 -output_matches_per_group 5 -
output_matchres_only false -enumerate_ligand_rotamers -
only_enumerate_non_match_redundant_ligand_rotamers -
out::file::output_virtual
```

The contents of the .pos file was as follows.

```
N_CST 1
1: 308
```

The contents of each constraint file varied slightly with the name of the ligand. An example constraint file for NapLuc-2-NMe_2 is as follows.

```
CST::BEGIN
NATIVE
  TEMPLATE::   ATOM_MAP: 1 atom_name: O1 C21 C8
  TEMPLATE::   ATOM_MAP: 1 residue3: NLA

  TEMPLATE::   ATOM_MAP: 2 atom_name: N CA C ,
  TEMPLATE::   ATOM_MAP: 2 residue1:  G
  TEMPLATE::   ATOM_MAP: 2 is_backbone

  CONSTRAINT:: distanceAB:   4.30   1.50  80.0  1     1
  CONSTRAINT::    angle_A:   135.3    5.0  10.0  360.  1
  CONSTRAINT::    angle_B:    43.6    5.0  10.0  360.  1
  CONSTRAINT::  torsion_A:    10.7    5.0  10.0  360.  1
  CONSTRAINT:: torsion_AB:  -160.7    5.0  10.0  360.  1
  CONSTRAINT::  torsion_B:  -134.1    5.0  10.0  360.  1

  ALGORITHM_INFO:: match
#     SECONDARY_MATCH: DOWNSTREAM
     CHI_STRATEGY:: CHI 1 EX_THREE_THIRD_STEP_STDDEVS
     CHI_STRATEGY:: CHI 2 EX_THREE_THIRD_STEP_STDDEVS
  ALGORITHM_INFO::END
CST::END
```

Matcher output .pdb files were used as input for the Rosetta CoupledMoves algorithm where mutagenesis, side chain conformational sampling and backbone minimization was undertaken to sculpt the binding pocket to accommodate the new ligand by alleviating any clashes between the ligand and the protein side chains for the CouLuc-2/NapLuc modeling. The Rosetta CoupledMoves application was called with the following script. (Ollikainen et al., 2015)

```
/path/to/Rosetta/coupled_moves.linuxgccrelease -jd2:ntrials 100 -
out:level 200 -s input.pdb -packing:resfile akaluc.resfile -database
~/Rosetta/main/database -mute protocols.backrub.BackrubMover -
extra_res_fa LCG_dimethyl.params -extra_res_fa NLA.params -ex1 -ex2 -
extrachi_cutoff 0 -nstruct 10 -coupled_moves::mc_kt 0.6 -
coupled_moves::ntrials 1000 -coupled_moves::initial_repack true -
coupled_moves::ligand_mode true -coupled_moves::fix_backbone false -
coupled_moves::bias_sampling true -coupled_moves::bump_check true -
coupled_moves::ligand_weight 2.0
```

For the CouLuc-3 modeling, the Matcher output files were used as input for the RosettaDesign algorithm where the active site pocket was also sculpted through side chain conformational sampling. The RosettaDesign application was called with the following script.

```
<path_to_Rosetta>/main/source/bin/rosetta_scripts.linuxgccrelease -
parser:script_vars ligand -out:nstruct 1 -extra_res_fa -
match:geometric_constraint_file -match:lig_name -jd2:ntrials 100 -
parser:protocol <xml file> -
match:scaffold_active_site_residues_for_geomcsts -in:file:native
4g36.pdb -database /main/database -s <input file from Matcher> -
run::preserve_header -unmute protocols.enzdes.EnzRepackMinimize -
enzdes::detect_design_interface -enzdes::cst_design -
enzdes::design_min_cycles 10 #The number of design-minimization
iterations -enzdes::minimize_ligand_torsions 5.0 -
enzdes::lig_packer_weight 1 -enzdes::cst_min -enzdes::bb_min -
enzdes:chi_min -packing::use_input_sc -packing::extrachi_cutoff 1 -
packing::ex1 -packing::ex2 -packing::soft_rep_design -linmem_ig 10 -
in:ignore_unrecognized_res -jd2:enzdes_out -nblist_autoupdate -
enzdes::bb_min_allowed_dev 0.05 -no_his_his_pairE
```

The contents of the RosettaDesign xml file were as follows.

```
<ROSETTASCRIPTS>

  <SCOREFXNS>
    <ScoreFunction name="ref15" weights="ref2015.wts"/>
      <ScoreFunction name="soft" weights="ref2015_soft.wts"/> #ref2015
recommended for protein repacking
      <ScoreFunction name="soft_cst" weights="ref2015_soft_cst.wts"/>
#soft-rep with constraints weights
      <ScoreFunction name="ref15_cst" weights="ref2015_cst.wts"/>
#hard-rep with constraints
  </SCOREFXNS>

  <RESIDUE_SELECTORS>
     <Index name="interface"
resnums="76,100,101,102,103,105,106,110,113,114,115,117,118,192,193,194
,195,196,197,198,199,210,213,214,219,220,221,229,233,234,236,237,238,23
9,243,246-
247,250,251,252,276,277,278,280,281,282,303,305,306,307,308,309,310,311
,312,313,316,317,320,326,327,328,239,330,332,332,333,334,335,336,338,33
9,340,343,344,345,348,349,350,351,352,353,354,355,414,423,426,428,429,4
31,433,434,435,437,438,439,440,441,442,445,446,449,469,470,471,498,499,
500,501,502"/>
    <ResidueName name="lig" residue_names="%%ligand%%"
residue_name3="%%ligand%%"/>
  </RESIDUE_SELECTORS>

  <FILTERS>
      <EnzScore name="allcst" score_type="cstE" whole_pose="1"
energy_cutoff="1000"/>
      </FILTERS>

  <SIMPLE_METRICS>
      <PerResidueEnergyMetric name="total_res_energies"
residue_selector="interface" scoretype="total_score" scorefxn="ref15"/>
      <PerResidueRMSDMetric name="res_RMSD"
residue_selector="interface" use_native="1" rmsd_type="rmsd_all"/>
  </SIMPLE_METRICS>

  <RESIDUE_LEVEL_TASK_OPERATIONS>
    <PreventRepackingRLT name="PreventRepacking"/>
  </RESIDUE_LEVEL_TASK_OPERATIONS>

  <TASKOPERATIONS>
      <DesignRestrictions name="notouch">
      <Action residue_selector="interface"
residue_level_operations="PreventRepacking"/>
            <Action residue_selector="lig"
residue_level_operations="PreventRepacking"/>
      </DesignRestrictions>

    <DesignRestrictions name="nomove">
            <Action residue_selector="lig"
residue_level_operations="PreventRepacking"/>
      </DesignRestrictions>
```

```
    <DetectProteinLigandInterface name="cuts_on" cut1="6" cut2="8"
cut3="10" cut4="12" design="1"/>
      <DetectProteinLigandInterface name="cuts_off" cut1="6" cut2="8"
cut3="10" cut4="12" design="0"/>

        <RestrictResiduesToRepacking name="pack_only"
residues="%%constrain%%"/>
      <ProteinLigandInterfaceUpweighter name="interface"
interface_weight="0.5"/>
  </TASKOPERATIONS>

  <MOVERS>

      <FastRelax name="relax" scorefxn="ref15" disable_design="1"
task_operations="notouch"/>

      <RunSimpleMetrics name="total_initial_energy"
metrics="total_res_energies" prefix="initial_"/>
      <RunSimpleMetrics name="RMSD" metrics="res_RMSD"/>
      <RunSimpleMetrics name="total_final_energy"
metrics="total_res_energies" prefix="final_"/>

            #Favor Native Residues
      <FavorNativeResidue name="favor_natives" bonus="0.75"/>

            #Add constraints to file
      <AddOrRemoveMatchCsts name="addcst" cst_instruction="add_new"/>
      <AddOrRemoveMatchCsts name="rmvcst" cst_instruction="remove"/>

            #Optimize the pose per the cst file
      <EnzRepackMinimize name="cstopt" scorefxn_repack="soft_cst"
scorefxn_minimize="ref15_cst" cst_opt="1" design="0" repack_only="1"
fix_catalytic="1" minimize_rb="1" minimize_bb="1" minimize_sc="1"
minimize_lig="1" min_in_stages="1" cycles="1"
task_operations="cuts_off"/>

            #Design and repacking around the catalytic residues; keep
the catalytic residues fixed in this instance.
      <EnzRepackMinimize name="dsgn" scorefxn_repack="soft_cst"
scorefxn_minimize="ref15_cst" design="1" repack_only="0"
fix_catalytic="1" minimize_rb="1" minimize_bb="1" minimize_sc="1"
minimize_lig="1" min_in_stages="1" cycles="1"
task_operations="cuts_on,nomove"/> #Added nomove to prevent ligand from
moving too much 2/17/22

            #Minimize scaffold and ligand after each design
      <EnzRepackMinimize name="min_cst" scorefxn_repack="soft_cst"
scorefxn_minimize="ref15_cst" design="0" repack_only="1"
fix_catalytic="0" minimize_rb="1" minimize_bb="1" minimize_sc="1"
minimize_lig="1" min_in_stages="1" cycles="1"
task_operations="cuts_on"/>
```

```
            #Perform a final repacking step.
      <EnzRepackMinimize name="repack" scorefxn_repack="soft"
scorefxn_minimize="ref15" design="0" repack_only="1" fix_catalytic="0"
minimize_rb="1" minimize_bb="0" minimize_sc="1" minimize_lig="0"
min_in_stages="0" cycles="1" task_operations="cuts_off"/>

            #Perform a final minimization step.
      <EnzRepackMinimize name="min" scorefxn_repack="soft"
scorefxn_minimize="ref15" design="0" repack_only="0" fix_catalytic="0"
minimize_rb="1" minimize_bb="1" minimize_sc="1" minimize_lig="0"
min_in_stages="0" cycles="1" task_operations="cuts_off"/>
  </MOVERS>

  <PROTOCOLS>
      <Add mover_name="total_initial_energy"/>
      <Add mover_name="favor_natives"/>
      <Add mover_name="addcst"/>
      <Add mover_name="cstopt"/>
      <Add mover_name="dsgn"/>
      <Add mover_name="min_cst"/>
      <Add mover_name="repack"/>
      <Add mover_name="min"/>
      <Add mover_name="relax"/>
      <Add mover_name="total_final_energy"/>
      <Add mover_name="RMSD"/>
      </PROTOCOLS>

</ROSETTASCRIPTS>
```

Models of library variants were built by making the requisite mutations during the RosettaDesign algorithm and during the CoupledMoves algorithm using a "resfile." A resfile is a Rosetta input file that indicates the locations and residue identities of the desired mutations. (Conway et al., 2014b; Nivón et al., 2013; Tyka et al., 2011). An example resfile to generate the mutations found in the Akaluc structure is as follows.

Akaluc resfile:

```
NATAA
start
36   A    PIKAA    A #T39A
45   A    PIKAA    Q #E48Q
48   A    PIKAA    V #I51V
65   A    PIKAA    R #K68R
83   A    PIKAA    S #L86S
131  A    PIKAA    R #Q134R
133  A    PIKAA    V #I136V
144  A    PIKAA    R #Q147R
172  A    PIKAA    S #G175S
```

```
221  A    PIKAA    Y #N229Y
223  A    PIKAA    N #I231N
286  A    PIKAA    C #F294C
287  A    PIKAA    L #F295L
300  A    PIKAA    S #N308S
302  A    PIKAA    R #H310R
324  A    PIKAA    R #H332R
339  A    PIKAA    N #S347N
341  A    PIKAA    V #I349V
342  A    PIKAA    M #L350M
349  A    PIKAA    R #D357R
353  A    PIKAA    S #A361S
369  A    PIKAA    V #D377V
448  A    PIKAA    G #S456G
#N463 missing density #N463Y
505  A    PIKAA    R #K524R
507  A    PIKAA    S #L526S
521 A    PIKAA    T #I540T
#G545D missing density
```


132

APPENDIX D

ADDITIONAL LIGAND PARAMS FILES FOR LUCIFERIN MODELS

The SLU.params file was as follows.

```
NAME SLU
IO_STRING SLU Z
TYPE LIGAND
AA UNK
ATOM  S3  S     X   1.62
ATOM  N8  Ntrp  X   -0.79
ATOM  C16 CNH2  X   0.81
ATOM  C14 CH1   X   0.21
ATOM  C8  aroC  X   0.08
ATOM  S2  S     X   -0.28
ATOM  C11 aroC  X   0.68
ATOM  C7  aroC  X   0.29
ATOM  N1  Nhis  X   -0.57
ATOM  C4  aroC  X   0.23
ATOM  C1  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C5  aroC  X   0.08
ATOM  C3  aroC  X   -0.15
ATOM  C6  aroC  X   0.04
ATOM  S1  S     X   -0.08
ATOM  H3  Haro  X   0.15
ATOM  O5  OH    X   -0.53
ATOM  H16 Hpol  X   0.45
ATOM  H2  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  N5  Nhis  X   -0.66
ATOM  H4  Haro  X   0.06
ATOM  O1  ONH2  X   -0.57
ATOM  H15 Hpol  X   0.42
ATOM  O2  OOC   X   -0.65
ATOM  O3  OOC   X   -0.65
ATOM  O8  OH    X   -0.46
ATOM  C21 CH2   X   0.28
ATOM  C19 CH1   X   0.28
ATOM  C17 CH1   X   0.28
ATOM  C18 CH1   X   0.28
ATOM  C20 CH1   X   0.63
ATOM  N6  Npro  X   -0.08
ATOM  C10 aroC  X   0.55
ATOM  N4  Nhis  X   -0.66
ATOM  C12 aroC  X   0.46
ATOM  C13 aroC  X   0.78
ATOM  N3  Nhis  X   -0.66
ATOM  C9  aroC  X   0.60
ATOM  N2  Nhis  X   -0.65
ATOM  C15 CH1   X   0.51
ATOM  N7  NH2O  X   -0.12
ATOM  H13 Hpol  X   0.40
ATOM  H14 Hpol  X   0.40
ATOM  H5  Haro  X   0.06
ATOM  H6  Haro  X   0.06
ATOM  O4  OH    X   -0.56
ATOM  H10 Hapo  X   0.00
```

```
ATOM  O7  OH    X   -0.68
ATOM  H18 Hpol  X   0.40
ATOM  H8  Hapo  X   0.00
ATOM  O6  OH    X   -0.68
ATOM  H17 Hpol  X   0.40
ATOM  H7  Hapo  X   0.00
ATOM  H9  Hapo  X   0.00
ATOM  H11 Hapo  X   0.00
ATOM  H12 Hapo  X   0.00
BOND_TYPE  C1   C2  4
BOND_TYPE  C1   C4  4
BOND_TYPE  C2   C5  4
BOND_TYPE  C3   C5  4
BOND_TYPE  C3   C6  4
BOND_TYPE  C4   C6  4
BOND_TYPE  C4   N1  1
BOND_TYPE  C5   O5  1
BOND_TYPE  C6   S1  1
BOND_TYPE  C7   C11 1
BOND_TYPE  C7   N1  2
BOND_TYPE  C7   S1  1
BOND_TYPE  C8   C14 1
BOND_TYPE  C8   S2  2
BOND_TYPE  C9   N2  2
BOND_TYPE  C9   N3  1
BOND_TYPE  C10  N4  1
BOND_TYPE  C10  N6  2
BOND_TYPE  C11  N5  2
BOND_TYPE  C11  S2  1
BOND_TYPE  C12  C13 1
BOND_TYPE  C12  C15 1
BOND_TYPE  C12  N4  2
BOND_TYPE  C13  N3  2
BOND_TYPE  C13  N6  1
BOND_TYPE  C14  C16 1
BOND_TYPE  C14  N5  1
BOND_TYPE  C15  N2  1
BOND_TYPE  C15  N7  1
BOND_TYPE  C16  N8  4
BOND_TYPE  C16  O1  2
BOND_TYPE  C17  C18 1
BOND_TYPE  C17  C19 1
BOND_TYPE  C17  O6  1
BOND_TYPE  C18  C20 1
BOND_TYPE  C18  O7  1
BOND_TYPE  C19  C21 1
BOND_TYPE  C19  O4  1
BOND_TYPE  C20  N6  1
BOND_TYPE  C20  O4  1
BOND_TYPE  C21  O8  1
BOND_TYPE  N8   S3  1
BOND_TYPE  O2   S3  2
BOND_TYPE  O3   S3  2
BOND_TYPE  O8   S3  1
BOND_TYPE  C1   H1  1
```

```
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C8   H4  1
BOND_TYPE  C9   H5  1
BOND_TYPE  C10  H6  1
BOND_TYPE  C17  H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C19  H9  1
BOND_TYPE  C20  H10 1
BOND_TYPE  C21  H11 1
BOND_TYPE  C21  H12 1
BOND_TYPE  N7   H13 1
BOND_TYPE  N7   H14 1
BOND_TYPE  N8   H15 1
BOND_TYPE  O5   H16 1
BOND_TYPE  O6   H17 1
BOND_TYPE  O7   H18 1
CHI 1  C2   C5   O5   H16
PROTON_CHI 1 SAMPLES 2 0 180 EXTRA 0
CHI 2  C19  C17  O6   H17
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 0
CHI 3  C17  C18  O7   H18
PROTON_CHI 3 SAMPLES 3 60 -60 180 EXTRA 0
CHI 4  S2   C11  C7   N1
CHI 5  N8   C16  C14  C8
CHI 6  O8   C21  C19  C17
CHI 7  C18  C20  N6   C10
CHI 8  S3   O8   C21  C19
CHI 9  O2   S3   N8   C16
CHI 10  N8   S3   O8   C21
NBR_ATOM  S3
NBR_RADIUS 15.331690
ICOOR_INTERNAL    S3     0.000000    0.000000    0.000000   S3    N8
C16
ICOOR_INTERNAL    N8     0.000000  180.000000    1.649468   S3    N8
C16
ICOOR_INTERNAL    C16    0.000000   58.301303    1.468440   N8    S3
C16
ICOOR_INTERNAL    C14 -179.026765   66.190937    1.467099   C16   N8
S3
ICOOR_INTERNAL    C8   150.001919   53.612295    1.382663   C14   C16
N8
ICOOR_INTERNAL    S2  -179.276669   69.544208    1.714009   C8    C14
C16
ICOOR_INTERNAL    C11    0.364755   89.734176    1.730152   S2    C8
C14
ICOOR_INTERNAL    C7   179.943935   57.759050    1.444385   C11   S2
C8
ICOOR_INTERNAL    N1    -0.837493   56.799721    1.318117   C7    C11
S2
ICOOR_INTERNAL    C4   179.869950   69.803179    1.384206   N1    C7
C11
ICOOR_INTERNAL    C1   179.911901   54.184816    1.409523   C4    N1
C7
```

```
ICOOR_INTERNAL    C2   179.936163   60.300737    1.398062   C1    C4
N1
ICOOR_INTERNAL    C5     0.003526   59.887010    1.387361   C2    C1
C4
ICOOR_INTERNAL    C3     0.007226   58.207192    1.394926   C5    C2
C1
ICOOR_INTERNAL    C6    -0.009962   62.553954    1.406198   C3    C5
C2
ICOOR_INTERNAL    S1  -179.977454   51.050316    1.720909   C6    C3
C5
ICOOR_INTERNAL    H3  -179.989158   57.933132    1.083926   C3    C5
C6
ICOOR_INTERNAL    O5   179.997951   62.706448    1.365157   C5    C2
C3
ICOOR_INTERNAL    H16  179.998826   70.905029    0.971545   O5    C5
C2
ICOOR_INTERNAL    H2   179.984819   60.095414    1.087046   C2    C1
C5
ICOOR_INTERNAL    H1   179.992009   60.262532    1.085520   C1    C4
C2
ICOOR_INTERNAL    N5   179.242791   66.371076    1.317856   C11   S2
C7
ICOOR_INTERNAL    H4   179.991860   55.230173    1.031940   C8    C14
S2
ICOOR_INTERNAL    O1   177.613201   56.617971    1.227139   C16   N8
C14
ICOOR_INTERNAL    H15 -179.995484   60.845852    0.984556   N8    S3
C16
ICOOR_INTERNAL    O2   -83.829697   72.117552    1.437687   S3    N8
C16
ICOOR_INTERNAL    O3  -110.978610   68.096527    1.445149   S3    N8
O2
ICOOR_INTERNAL    O8  -129.796645   67.882755    1.508902   S3    N8
O3
ICOOR_INTERNAL    C21   47.972090   59.846860    1.411589   O8    S3
N8
ICOOR_INTERNAL    C19  157.664803   70.737827    1.511020   C21   O8
S3
ICOOR_INTERNAL    C17 -162.089472   70.934687    1.469290   C19   C21
O8
ICOOR_INTERNAL    C18 -144.250635   77.197183    1.465518   C17   C19
C21
ICOOR_INTERNAL    C20   31.359911   77.525336    1.475945   C18   C17
C19
ICOOR_INTERNAL    N6    91.982589   65.434427    1.445330   C20   C18
C17
ICOOR_INTERNAL    C10  -93.612286   52.726401    1.352130   N6    C20
C18
ICOOR_INTERNAL    N4  -179.125529   69.963424    1.324697   C10   N6
C20
ICOOR_INTERNAL    C12    0.434724   70.201937    1.328569   N4    C10
N6
ICOOR_INTERNAL    C13   -0.108966   73.543980    1.420687   C12   N4
C10
```

```
ICOOR_INTERNAL    N3   179.727076   60.631876    1.342184   C13   C12
N4
ICOOR_INTERNAL    C9     0.605797   59.617650    1.325985   N3    C13
C12
ICOOR_INTERNAL    N2    -0.843694   57.615869    1.331578   C9    N3
C13
ICOOR_INTERNAL    C15    0.543846   58.116997    1.342047   N2    C9
N3
ICOOR_INTERNAL    N7  -179.885925   60.505121    1.449333   C15   N2
C9
ICOOR_INTERNAL    H13  179.857903   59.996676    0.984436   N7    C15
N2
ICOOR_INTERNAL    H14 -179.996362   60.003205    0.984528   N7    C15
H13
ICOOR_INTERNAL    H5   179.991347   61.186872    1.031972   C9    N3
N2
ICOOR_INTERNAL    H6   179.994533   55.014941    1.032043   C10   N6
N4
ICOOR_INTERNAL    O4  -122.220593   71.733603    1.413878   C20   C18
N6
ICOOR_INTERNAL    H10 -121.131406   73.232551    1.069922   C20   C18
O4
ICOOR_INTERNAL    O7  -116.507140   66.309831    1.374491   C18   C17
C20
ICOOR_INTERNAL    H18  179.997609   70.527483    0.969954   O7    C18
C17
ICOOR_INTERNAL    H8  -120.201693   66.638819    1.112365   C18   C17
O7
ICOOR_INTERNAL    O6  -122.102731   69.950997    1.380152   C17   C19
C18
ICOOR_INTERNAL    H17  -64.962631   70.528340    0.970025   O6    C17
C19
ICOOR_INTERNAL    H7  -117.021009   64.837781    1.070175   C17   C19
O6
ICOOR_INTERNAL    H9   119.990690   70.467539    1.070000   C19   C21
C17
ICOOR_INTERNAL    H11 -119.973622   70.474860    1.069935   C21   O8
C19
ICOOR_INTERNAL    H12 -119.929165   70.409662    1.070081   C21   O8
H11
```

The CouLuc-3-NMe_2 .params file for the Z isomer is below.

```
NAME LCA
IO_STRING LCA Z
TYPE LIGAND
AA UNK
ATOM  C21 CNH2  X   0.89
ATOM  C8  aroC  X   0.14
ATOM  C4  aroC  X   -0.11
ATOM  S1  S     X   -0.08
ATOM  C9  aroC  X   0.33
ATOM  C18 aroC  X   -0.10
ATOM  C20 aroC  X   -0.15
ATOM  C19 aroC  X   -0.15
```

```
ATOM  C15 aroC  X   0.08
ATOM  C10 aroC  X   -0.15
ATOM  C13 aroC  X   -0.17
ATOM  C5  aroC  X   0.03
ATOM  C1  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C6  aroC  X   0.10
ATOM  C3  aroC  X   -0.15
ATOM  C7  aroC  X   0.08
ATOM  O4  Oaro  X   -0.16
ATOM  H3  Haro  X   0.15
ATOM  N8  Nhis  X   -0.84
ATOM  C26 CH3   X   0.37
ATOM  H15 Hapo  X   0.00
ATOM  H16 Hapo  X   0.00
ATOM  H17 Hapo  X   0.00
ATOM  C27 CH3   X   0.37
ATOM  H18 Hapo  X   0.00
ATOM  H19 Hapo  X   0.00
ATOM  H20 Hapo  X   0.00
ATOM  H2  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  C29 CH1   X   1.16
ATOM  F1  F     X   -0.34
ATOM  F2  F     X   -0.34
ATOM  F3  F     X   -0.34
ATOM  H5  Haro  X   0.15
ATOM  H9  Haro  X   0.15
ATOM  H10 Haro  X   0.15
ATOM  H8  Haro  X   0.15
ATOM  N1  Nhis  X   -0.57
ATOM  H4  Haro  X   0.15
ATOM  N7  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O2  OOC   X   -0.65
ATOM  O3  OOC   X   -0.65
ATOM  O8  OH    X   -0.46
ATOM  C28 CH2   X   0.28
ATOM  C24 CH1   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C23 CH1   X   0.28
ATOM  C25 CH1   X   0.63
ATOM  N5  Npro  X   -0.08
ATOM  C12 aroC  X   0.55
ATOM  N4  Nhis  X   -0.66
ATOM  C14 aroC  X   0.46
ATOM  C16 aroC  X   0.78
ATOM  N3  Nhis  X   -0.66
ATOM  C11 aroC  X   0.60
ATOM  N2  Nhis  X   -0.65
ATOM  C17 CH1   X   0.51
ATOM  N6  NH2O  X   -0.12
ATOM  H23 Hpol  X   0.40
ATOM  H24 Hpol  X   0.40
ATOM  H6  Haro  X   0.06
```

```
ATOM  H7  Haro  X   0.06
ATOM  O5  OH    X   -0.56
ATOM  H14 Hapo  X   0.00
ATOM  O7  OH    X   -0.68
ATOM  H27 Hpol  X   0.40
ATOM  H12 Hapo  X   0.00
ATOM  O6  OH    X   -0.68
ATOM  H26 Hpol  X   0.40
ATOM  H11 Hapo  X   0.00
ATOM  H13 Hapo  X   0.00
ATOM  H21 Hapo  X   0.00
ATOM  H22 Hapo  X   0.00
ATOM  H25 Hpol  X   0.42
ATOM  O1  ONH2  X   -0.57
BOND_TYPE  C1   C2  4
BOND_TYPE  C1   C5  4
BOND_TYPE  C2   C6  4
BOND_TYPE  C3   C6  4
BOND_TYPE  C3   C7  4
BOND_TYPE  C4   C8  2
BOND_TYPE  C4   S1  1
BOND_TYPE  C5   C7  4
BOND_TYPE  C5   C13 1
BOND_TYPE  C6   N8  1
BOND_TYPE  C7   O4  1
BOND_TYPE  C8   C21 1
BOND_TYPE  C8   N1  1
BOND_TYPE  C9   C18 1
BOND_TYPE  C9   N1  2
BOND_TYPE  C9   S1  1
BOND_TYPE  C10  C13 2
BOND_TYPE  C10  C15 1
BOND_TYPE  C11  N2  2
BOND_TYPE  C11  N3  1
BOND_TYPE  C12  N4  1
BOND_TYPE  C12  N5  2
BOND_TYPE  C13  C29 1
BOND_TYPE  C14  C16 1
BOND_TYPE  C14  C17 1
BOND_TYPE  C14  N4  2
BOND_TYPE  C15  C19 2
BOND_TYPE  C15  O4  1
BOND_TYPE  C16  N3  2
BOND_TYPE  C16  N5  1
BOND_TYPE  C17  N2  1
BOND_TYPE  C17  N6  1
BOND_TYPE  C18  C20 2
BOND_TYPE  C19  C20 1
BOND_TYPE  C21  N7  4
BOND_TYPE  C21  O1  2
BOND_TYPE  C22  C23 1
BOND_TYPE  C22  C24 1
BOND_TYPE  C22  O6  1
BOND_TYPE  C23  C25 1
BOND_TYPE  C23  O7  1
```

```
BOND_TYPE  C24  C28 1
BOND_TYPE  C24  O5  1
BOND_TYPE  C25  N5  1
BOND_TYPE  C25  O5  1
BOND_TYPE  C26  N8  1
BOND_TYPE  C27  N8  1
BOND_TYPE  C28  O8  1
BOND_TYPE  C29  F1  1
BOND_TYPE  C29  F2  1
BOND_TYPE  C29  F3  1
BOND_TYPE  N7   S2  1
BOND_TYPE  O2   S2  2
BOND_TYPE  O3   S2  2
BOND_TYPE  O8   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C4   H4  1
BOND_TYPE  C10  H5  1
BOND_TYPE  C11  H6  1
BOND_TYPE  C12  H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C19  H9  1
BOND_TYPE  C20  H10 1
BOND_TYPE  C22  H11 1
BOND_TYPE  C23  H12 1
BOND_TYPE  C24  H13 1
BOND_TYPE  C25  H14 1
BOND_TYPE  C26  H15 1
BOND_TYPE  C26  H16 1
BOND_TYPE  C26  H17 1
BOND_TYPE  C27  H18 1
BOND_TYPE  C27  H19 1
BOND_TYPE  C27  H20 1
BOND_TYPE  C28  H21 1
BOND_TYPE  C28  H22 1
BOND_TYPE  N6   H23 1
BOND_TYPE  N6   H24 1
BOND_TYPE  N7   H25 1
BOND_TYPE  O6   H26 1
BOND_TYPE  O7   H27 1
CHI 1  C24  C22  O6   H26
PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 0
CHI 2  C22  C23  O7   H27
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 0
CHI 3  C2   C6   N8   C26
CHI 4  N7   C21  C8   C4
CHI 5  S1   C9   C18  C20
CHI 6  C10  C13  C29  F1
CHI 7  C18  C20  C19  C15
CHI 8  O8   C28  C24  C22
CHI 9  C23  C25  N5   C12
CHI 10  S2   O8   C28  C24
CHI 11  C21  N7   S2   O2
CHI 12  N7   S2   O8   C28
```

```
NBR_ATOM  C21
NBR_RADIUS 17.590121
ICOOR_INTERNAL    C21    0.000000    0.000000    0.000000   C21   C8    C4
ICOOR_INTERNAL    C8     0.000000  179.999999    1.467129   C21   C8    C4
ICOOR_INTERNAL    C4     0.000001   53.617032    1.382669   C8    C21   C4
ICOOR_INTERNAL    S1  -179.281387   69.551912    1.714111   C4    C8    C21
ICOOR_INTERNAL    C9     0.368538   89.731913    1.730053   S1    C4    C8
ICOOR_INTERNAL    C18 -179.983428   55.061880    1.476720   C9    S1    C4
ICOOR_INTERNAL    C20  179.107511   57.154412    1.342614   C18   C9    S1
ICOOR_INTERNAL    C19  179.963866   59.541585    1.485000   C20   C18   C9
ICOOR_INTERNAL    C15  179.999695   54.513145    1.349071   C19   C20   C18
ICOOR_INTERNAL    C10   -0.053775   56.133742    1.491760   C15   C19   C20
ICOOR_INTERNAL    C13 -179.979235   59.844397    1.351898   C10   C15   C19
ICOOR_INTERNAL    C5    -0.016465   61.367697    1.498679   C13   C10   C15
ICOOR_INTERNAL    C1  -179.978742   57.380727    1.404750   C5    C13   C10
ICOOR_INTERNAL    C2   179.978603   59.280050    1.396556   C1    C5    C13
ICOOR_INTERNAL    C6    -0.011294   58.773291    1.412211   C2    C1    C5
ICOOR_INTERNAL    C3     0.011612   62.377933    1.412764   C6    C2    C1
ICOOR_INTERNAL    C7    -0.002535   58.740853    1.398770   C3    C6    C2
ICOOR_INTERNAL    O4   179.994190   61.003563    1.354369   C7    C3    C6
ICOOR_INTERNAL    H3   179.996756   58.715470    1.080834   C3    C6    C7
ICOOR_INTERNAL    N8  -179.978952   58.833041    1.444598   C6    C2    C3
ICOOR_INTERNAL    C26   -0.005874   59.369540    1.456963   N8    C6    C2
ICOOR_INTERNAL    H15 -179.999668   70.119058    1.110614   C26   N8    C6
ICOOR_INTERNAL    H16 -120.878814   70.148362    1.110679   C26   N8    H15
ICOOR_INTERNAL    H17 -119.538468   67.587037    1.110219   C26   N8    H16
ICOOR_INTERNAL    C27  179.991082   59.353693    1.457042   N8    C6    C26
ICOOR_INTERNAL    H18  179.997383   70.129063    1.110567   C27   N8    C6
ICOOR_INTERNAL    H19 -119.544038   67.600229    1.110291   C27   N8    H18
ICOOR_INTERNAL    H20 -119.542316   70.127139    1.110547   C27   N8    H19
ICOOR_INTERNAL    H2  -179.980340   62.733103    1.080662   C2    C1    C6
ICOOR_INTERNAL    H1  -179.996344   58.605071    1.080428   C1    C5    C2
ICOOR_INTERNAL    C29 -179.995527   59.589585    1.520781   C13   C10   C5
ICOOR_INTERNAL    F1   179.998863   67.610783    1.385073   C29   C13   C10
ICOOR_INTERNAL    F2   120.037494   70.475588    1.383039   C29   C13   F1
ICOOR_INTERNAL    F3   119.916915   70.471215    1.382932   C29   C13   F2
ICOOR_INTERNAL    H5   179.977690   59.689688    1.080359   C10   C15   C13
ICOOR_INTERNAL    H9   179.992072   62.715893    1.087842   C19   C20   C15
ICOOR_INTERNAL    H10 -179.998350   61.078692    1.080735   C20   C18   C19
ICOOR_INTERNAL    H8  -179.998975   62.397857    1.087144   C18   C9    C20
ICOOR_INTERNAL    N1   179.169058   66.367720    1.317939   C9    S1    C18
ICOOR_INTERNAL    H4   178.050628   50.571171    1.080506   C4    C8    S1
ICOOR_INTERNAL    N7   179.999349   66.189768    1.468382   C21   C8    C4
ICOOR_INTERNAL    S2  -179.028528   58.301695    1.649480   N7    C21   C8
ICOOR_INTERNAL    O2   -83.828030   72.115214    1.437728   S2    N7    C21
ICOOR_INTERNAL    O3  -110.979907   68.095120    1.445125   S2    N7    O2
ICOOR_INTERNAL    O8  -129.796485   67.884564    1.508898   S2    N7    O3
ICOOR_INTERNAL    C28   47.972804   59.847207    1.411561   O8    S2    N7
ICOOR_INTERNAL    C24  157.660576   70.738957    1.511080   C28   O8    S2
ICOOR_INTERNAL    C22 -162.092351   70.936507    1.469262   C24   C28   O8
ICOOR_INTERNAL    C23 -144.249833   77.197932    1.465575   C22   C24   C28
ICOOR_INTERNAL    C25   31.366047   77.527376    1.475916   C23   C22   C24
ICOOR_INTERNAL    N5    91.982349   65.429978    1.445332   C25   C23   C22
ICOOR_INTERNAL    C12  -93.611780   52.728629    1.352156   N5    C25   C23
ICOOR_INTERNAL    N4  -179.129471   69.965230    1.324715   C12   N5    C25
```

```
ICOOR_INTERNAL    C14    0.433861   70.203038    1.328601   N4    C12   N5
ICOOR_INTERNAL    C16   -0.110422   73.543260    1.420679   C14   N4    C12
ICOOR_INTERNAL    N3   179.726204   60.629497    1.342216   C16   C14   N4
ICOOR_INTERNAL    C11    0.608176   59.624923    1.326020   N3    C16   C14
ICOOR_INTERNAL    N2    -0.848435   57.610071    1.331554   C11   N3    C16
ICOOR_INTERNAL    C17    0.552053   58.120141    1.342096   N2    C11   N3
ICOOR_INTERNAL    N6  -179.890730   60.506917    1.449335   C17   N2    C11
ICOOR_INTERNAL    H23  179.859865   60.001164    0.984524   N6    C17   N2
ICOOR_INTERNAL    H24  179.991250   60.000412    0.984519   N6    C17   H23
ICOOR_INTERNAL    H6  -179.999269   61.192473    1.031974   C11   N3    N2
ICOOR_INTERNAL    H7   179.997573   55.017697    1.031995   C12   N5    N4
ICOOR_INTERNAL    O5  -122.223815   71.733653    1.413900   C25   C23   N5
ICOOR_INTERNAL    H14 -121.130496   73.235732    1.069989   C25   C23   O5
ICOOR_INTERNAL    O7  -116.511459   66.309694    1.374530   C23   C22   C25
ICOOR_INTERNAL    H27 -179.997776   70.530952    0.969950   O7    C23   C22
ICOOR_INTERNAL    H12 -118.778660   69.018066    1.070015   C23   C22   O7
ICOOR_INTERNAL    O6  -122.097254   69.957154    1.380198   C22   C24   C23
ICOOR_INTERNAL    H26  -64.966530   70.526952    0.970044   O6    C22   C24
ICOOR_INTERNAL    H11 -117.021961   64.834299    1.070064   C22   C24   O6
ICOOR_INTERNAL    H13  120.000334   70.466931    1.069977   C24   C28   C22
ICOOR_INTERNAL    H21 -119.967044   70.476305    1.069971   C28   O8    C24
ICOOR_INTERNAL    H22 -119.934519   70.408073    1.070003   C28   O8    H21
ICOOR_INTERNAL    H25 -179.999258   60.849549    0.984516   N7    C21   S2
ICOOR_INTERNAL    O1  -177.630746   57.240813    1.227206   C21   C8    N7
PDB_ROTAMERS LCAa_conformers.pdb
```

The CouLuc-3-NMe_2 .params file for the E isomer is below.

```
NAME LCA
IO_STRING LCA Z
TYPE LIGAND
AA UNK
ATOM  C21 CNH2  X   0.89
ATOM  C8  aroC  X   0.14
ATOM  C4  aroC  X   -0.11
ATOM  S1  S     X   -0.08
ATOM  C9  aroC  X   0.33
ATOM  C18 aroC  X   -0.10
ATOM  C20 aroC  X   -0.15
ATOM  C19 aroC  X   -0.15
ATOM  C15 aroC  X   0.08
ATOM  C10 aroC  X   -0.15
ATOM  C13 aroC  X   -0.17
ATOM  C5  aroC  X   0.03
ATOM  C1  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C6  aroC  X   0.10
ATOM  C3  aroC  X   -0.15
ATOM  C7  aroC  X   0.08
ATOM  O4  Oaro  X   -0.16
ATOM  H3  Haro  X   0.15
ATOM  N8  Nhis  X   -0.84
ATOM  C26 CH3   X   0.37
ATOM  H15 Hapo  X   0.00
```

```
ATOM  H16 Hapo  X   0.00
ATOM  H17 Hapo  X   0.00
ATOM  C27 CH3   X   0.37
ATOM  H18 Hapo  X   0.00
ATOM  H19 Hapo  X   0.00
ATOM  H20 Hapo  X   0.00
ATOM  H2  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  C29 CH1   X   1.16
ATOM  F1  F     X   -0.34
ATOM  F2  F     X   -0.34
ATOM  F3  F     X   -0.34
ATOM  H5  Haro  X   0.15
ATOM  H9  Haro  X   0.15
ATOM  H10 Haro  X   0.15
ATOM  H8  Haro  X   0.15
ATOM  N1  Nhis  X   -0.57
ATOM  H4  Haro  X   0.15
ATOM  N7  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O2  OOC   X   -0.65
ATOM  O3  OOC   X   -0.65
ATOM  O8  OH    X   -0.46
ATOM  C28 CH2   X   0.28
ATOM  C24 CH1   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C23 CH1   X   0.28
ATOM  C25 CH1   X   0.63
ATOM  N5  Npro  X   -0.08
ATOM  C12 aroC  X   0.55
ATOM  N4  Nhis  X   -0.66
ATOM  C14 aroC  X   0.46
ATOM  C16 aroC  X   0.78
ATOM  N3  Nhis  X   -0.66
ATOM  C11 aroC  X   0.60
ATOM  N2  Nhis  X   -0.65
ATOM  C17 CH1   X   0.51
ATOM  N6  NH2O  X   -0.12
ATOM  H23 Hpol  X   0.40
ATOM  H24 Hpol  X   0.40
ATOM  H6  Haro  X   0.06
ATOM  H7  Haro  X   0.06
ATOM  O5  OH    X   -0.56
ATOM  H14 Hapo  X   0.00
ATOM  O7  OH    X   -0.68
ATOM  H27 Hpol  X   0.40
ATOM  H12 Hapo  X   0.00
ATOM  O6  OH    X   -0.68
ATOM  H26 Hpol  X   0.40
ATOM  H11 Hapo  X   0.00
ATOM  H13 Hapo  X   0.00
ATOM  H21 Hapo  X   0.00
ATOM  H22 Hapo  X   0.00
ATOM  H25 Hpol  X   0.42
ATOM  O1  ONH2  X   -0.57
```

```
BOND_TYPE  C1   C2  4
BOND_TYPE  C1   C5  4
BOND_TYPE  C2   C6  4
BOND_TYPE  C3   C6  4
BOND_TYPE  C3   C7  4
BOND_TYPE  C4   C8  2
BOND_TYPE  C4   S1  1
BOND_TYPE  C5   C7  4
BOND_TYPE  C5   C13 1
BOND_TYPE  C6   N8  1
BOND_TYPE  C7   O4  1
BOND_TYPE  C8   C21 1
BOND_TYPE  C8   N1  1
BOND_TYPE  C9   C18 1
BOND_TYPE  C9   N1  2
BOND_TYPE  C9   S1  1
BOND_TYPE  C10  C13 2
BOND_TYPE  C10  C15 1
BOND_TYPE  C11  N2  2
BOND_TYPE  C11  N3  1
BOND_TYPE  C12  N4  1
BOND_TYPE  C12  N5  2
BOND_TYPE  C13  C29 1
BOND_TYPE  C14  C16 1
BOND_TYPE  C14  C17 1
BOND_TYPE  C14  N4  2
BOND_TYPE  C15  C19 2
BOND_TYPE  C15  O4  1
BOND_TYPE  C16  N3  2
BOND_TYPE  C16  N5  1
BOND_TYPE  C17  N2  1
BOND_TYPE  C17  N6  1
BOND_TYPE  C18  C20 2
BOND_TYPE  C19  C20 1
BOND_TYPE  C21  N7  4
BOND_TYPE  C21  O1  2
BOND_TYPE  C22  C23 1
BOND_TYPE  C22  C24 1
BOND_TYPE  C22  O6  1
BOND_TYPE  C23  C25 1
BOND_TYPE  C23  O7  1
BOND_TYPE  C24  C28 1
BOND_TYPE  C24  O5  1
BOND_TYPE  C25  N5  1
BOND_TYPE  C25  O5  1
BOND_TYPE  C26  N8  1
BOND_TYPE  C27  N8  1
BOND_TYPE  C28  O8  1
BOND_TYPE  C29  F1  1
BOND_TYPE  C29  F2  1
BOND_TYPE  C29  F3  1
BOND_TYPE  N7   S2  1
BOND_TYPE  O2   S2  2
BOND_TYPE  O3   S2  2
BOND_TYPE  O8   S2  1
```

```
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C4   H4  1
BOND_TYPE  C10  H5  1
BOND_TYPE  C11  H6  1
BOND_TYPE  C12  H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C19  H9  1
BOND_TYPE  C20  H10 1
BOND_TYPE  C22  H11 1
BOND_TYPE  C23  H12 1
BOND_TYPE  C24  H13 1
BOND_TYPE  C25  H14 1
BOND_TYPE  C26  H15 1
BOND_TYPE  C26  H16 1
BOND_TYPE  C26  H17 1
BOND_TYPE  C27  H18 1
BOND_TYPE  C27  H19 1
BOND_TYPE  C27  H20 1
BOND_TYPE  C28  H21 1
BOND_TYPE  C28  H22 1
BOND_TYPE  N6   H23 1
BOND_TYPE  N6   H24 1
BOND_TYPE  N7   H25 1
BOND_TYPE  O6   H26 1
BOND_TYPE  O7   H27 1
CHI 1  C24  C22  O6   H26
PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 0
CHI 2  C22  C23  O7   H27
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 0
CHI 3  C2   C6   N8   C26
CHI 4  N7   C21  C8   C4
CHI 5  S1   C9   C18  C20
CHI 6  C10  C13  C29  F1
CHI 7  C18  C20  C19  C15
CHI 8  O8   C28  C24  C22
CHI 9  C23  C25  N5   C12
CHI 10  S2   O8   C28  C24
CHI 11  C21  N7   S2   O2
CHI 12  N7   S2   O8   C28
NBR_ATOM  C21
NBR_RADIUS 17.643938
ICOOR_INTERNAL    C21    0.000000    0.000000    0.000000   C21   C8    C4
ICOOR_INTERNAL    C8     0.000000  180.000000    1.467099   C21   C8    C4
ICOOR_INTERNAL    C4     0.000001   53.613584    1.382650   C8    C21   C4
ICOOR_INTERNAL    S1  -179.277244   69.545521    1.714033   C4    C8    C21
ICOOR_INTERNAL    C9     0.366000   89.733106    1.730169   S1    C4    C8
ICOOR_INTERNAL    C18  178.480672   53.956919    1.472190   C9    S1    C4
ICOOR_INTERNAL    C20 -179.253693   55.590211    1.341657   C18   C9    S1
ICOOR_INTERNAL    C19 -178.860703   61.081220    1.480117   C20   C18   C9
ICOOR_INTERNAL    C15  179.997551   56.075757    1.342682   C19   C20   C18
ICOOR_INTERNAL    C10 -179.356063   60.754970    1.477869   C15   C19   C20
ICOOR_INTERNAL    C13  179.033457   58.995367    1.347258   C10   C15   C19
ICOOR_INTERNAL    C5     0.096313   62.200717    1.513467   C13   C10   C15
```

```
ICOOR_INTERNAL    C1  -179.627334   56.000846    1.402666   C5    C13   C10
ICOOR_INTERNAL    C2  -179.917507   57.810018    1.397502   C1    C5    C13
ICOOR_INTERNAL    C6     0.093158   59.285238    1.414135   C2    C1    C5
ICOOR_INTERNAL    C3    -0.127674   62.800534    1.416784   C6    C2    C1
ICOOR_INTERNAL    C7    -0.012295   58.232499    1.401153   C3    C6    C2
ICOOR_INTERNAL    O4  -179.599495   61.355171    1.358145   C7    C3    C6
ICOOR_INTERNAL    H3  -178.593087   58.090090    1.078016   C3    C6    C7
ICOOR_INTERNAL    N8  -179.971338   58.616652    1.449221   C6    C2    C3
ICOOR_INTERNAL    C26   -0.001984   57.458357    1.457852   N8    C6    C2
ICOOR_INTERNAL    H15  179.999884   65.780358    1.102888   C26   N8    C6
ICOOR_INTERNAL    H16 -122.635202   70.083946    1.110682   C26   N8    H15
ICOOR_INTERNAL    H17 -119.309917   70.640070    1.110827   C26   N8    H16
ICOOR_INTERNAL    C27  179.959626   57.499291    1.457905   N8    C6    C26
ICOOR_INTERNAL    H18  179.995176   70.637767    1.110918   C27   N8    C6
ICOOR_INTERNAL    H19 -118.086256   65.797498    1.102983   C27   N8    H18
ICOOR_INTERNAL    H20 -122.600548   70.079213    1.110695   C27   N8    H19
ICOOR_INTERNAL    H2   178.693040   63.235478    1.078324   C2    C1    C6
ICOOR_INTERNAL    H1   179.717317   57.022015    1.070466   C1    C5    C2
ICOOR_INTERNAL    C29 -179.946573   64.677245    1.528239   C13   C10   C5
ICOOR_INTERNAL    F1   179.945748   64.850453    1.374843   C29   C13   C10
ICOOR_INTERNAL    F2   120.797426   70.974799    1.384186   C29   C13   F1
ICOOR_INTERNAL    F3   118.349656   70.966161    1.384147   C29   C13   F2
ICOOR_INTERNAL    H5  -179.768550   61.078044    1.088585   C10   C15   C13
ICOOR_INTERNAL    H9  -179.310715   62.326928    1.088433   C19   C20   C15
ICOOR_INTERNAL    H10  179.101411   59.204346    1.085987   C20   C18   C19
ICOOR_INTERNAL    H8  -178.976872   63.534159    1.087683   C18   C9    C20
ICOOR_INTERNAL    N1  -179.293981   66.372571    1.317944   C9    S1    C18
ICOOR_INTERNAL    H4   178.052410   50.569171    1.080507   C4    C8    S1
ICOOR_INTERNAL    N7   179.999853   66.190937    1.468440   C21   C8    C4
ICOOR_INTERNAL    S2  -179.026765   58.301303    1.649468   N7    C21   C8
ICOOR_INTERNAL    O2   -83.829697   72.117552    1.437687   S2    N7    C21
ICOOR_INTERNAL    O3  -110.978610   68.096527    1.445149   S2    N7    O2
ICOOR_INTERNAL    O8  -129.796645   67.882755    1.508902   S2    N7    O3
ICOOR_INTERNAL    C28   47.972090   59.846860    1.411589   O8    S2    N7
ICOOR_INTERNAL    C24  157.664803   70.737827    1.511020   C28   O8    S2
ICOOR_INTERNAL    C22 -162.089472   70.934687    1.469290   C24   C28   O8
ICOOR_INTERNAL    C23 -144.250635   77.197183    1.465518   C22   C24   C28
ICOOR_INTERNAL    C25   31.359911   77.525336    1.475945   C23   C22   C24
ICOOR_INTERNAL    N5    91.982589   65.434427    1.445330   C25   C23   C22
ICOOR_INTERNAL    C12  -93.612286   52.726401    1.352130   N5    C25   C23
ICOOR_INTERNAL    N4  -179.125529   69.963424    1.324697   C12   N5    C25
ICOOR_INTERNAL    C14    0.434724   70.201937    1.328569   N4    C12   N5
ICOOR_INTERNAL    C16   -0.108966   73.543980    1.420687   C14   N4    C12
ICOOR_INTERNAL    N3   179.727076   60.631876    1.342184   C16   C14   N4
ICOOR_INTERNAL    C11    0.605797   59.617650    1.325985   N3    C16   C14
ICOOR_INTERNAL    N2    -0.843694   57.615869    1.331578   C11   N3    C16
ICOOR_INTERNAL    C17    0.543846   58.116997    1.342047   N2    C11   N3
ICOOR_INTERNAL    N6  -179.885925   60.505121    1.449333   C17   N2    C11
ICOOR_INTERNAL    H23  179.857903   59.996676    0.984436   N6    C17   N2
ICOOR_INTERNAL    H24 -179.996362   60.003205    0.984528   N6    C17   H23
ICOOR_INTERNAL    H6   179.991347   61.186872    1.031972   C11   N3    N2
ICOOR_INTERNAL    H7   179.994533   55.014941    1.032043   C12   N5    N4
ICOOR_INTERNAL    O5  -122.220593   71.733603    1.413878   C25   C23   N5
ICOOR_INTERNAL    H14 -121.131406   73.232551    1.069922   C25   C23   O5
ICOOR_INTERNAL    O7  -116.507140   66.309831    1.374491   C23   C22   C25
```

```
ICOOR_INTERNAL    H27  179.997609   70.527483    0.969954   O7    C23   C22
ICOOR_INTERNAL    H12 -120.201693   66.638819    1.112365   C23   C22   O7
ICOOR_INTERNAL    O6  -122.102731   69.950997    1.380152   C22   C24   C23
ICOOR_INTERNAL    H26  -64.962631   70.528340    0.970025   O6    C22   C24
ICOOR_INTERNAL    H11 -117.021009   64.837781    1.070175   C22   C24   O6
ICOOR_INTERNAL    H13  119.990690   70.467539    1.070000   C24   C28   C22
ICOOR_INTERNAL    H21 -119.973622   70.474860    1.069935   C28   O8    C24
ICOOR_INTERNAL    H22 -119.929165   70.409662    1.070081   C28   O8    H21
ICOOR_INTERNAL    H25  179.995484   60.852845    0.984556   N7    C21   S2
ICOOR_INTERNAL    O1  -177.629925   57.236261    1.227139   C21   C8    N7
PDB_ROTAMERS LCAb_conformers.pdb
```

The CouLuc-3-OH .params file for the Z isomer is below.

```
NAME LCB
IO_STRING LCB Z
TYPE LIGAND
AA UNK
ATOM  C21 CNH2  X   0.89
ATOM  C8  aroC  X   0.14
ATOM  C4  aroC  X   -0.11
ATOM  S1  S     X   -0.08
ATOM  C9  aroC  X   0.33
ATOM  C18 aroC  X   -0.10
ATOM  C20 aroC  X   -0.15
ATOM  C19 aroC  X   -0.15
ATOM  C15 aroC  X   0.08
ATOM  C10 aroC  X   -0.15
ATOM  C13 aroC  X   -0.17
ATOM  C5  aroC  X   0.03
ATOM  C1  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C7  aroC  X   0.08
ATOM  C3  aroC  X   -0.15
ATOM  C6  aroC  X   0.08
ATOM  O4  Oaro  X   -0.16
ATOM  H3  Haro  X   0.15
ATOM  O6  OH    X   -0.53
ATOM  H20 Hpol  X   0.45
ATOM  H2  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  C27 CH1   X   1.16
ATOM  F1  F     X   -0.34
ATOM  F2  F     X   -0.34
ATOM  F3  F     X   -0.34
ATOM  H5  Haro  X   0.15
ATOM  H9  Haro  X   0.15
ATOM  H10 Haro  X   0.15
ATOM  H8  Haro  X   0.15
ATOM  N1  Nhis  X   -0.57
ATOM  H4  Haro  X   0.15
ATOM  N7  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O2  OOC   X   -0.65
```

```
ATOM  O3  OOC   X   -0.65
ATOM  O9  OH    X   -0.46
ATOM  C26 CH2   X   0.28
ATOM  C24 CH1   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C23 CH1   X   0.28
ATOM  C25 CH1   X   0.63
ATOM  N5  Npro  X   -0.08
ATOM  C12 aroC  X   0.55
ATOM  N4  Nhis  X   -0.66
ATOM  C14 aroC  X   0.46
ATOM  C16 aroC  X   0.78
ATOM  N3  Nhis  X   -0.66
ATOM  C11 aroC  X   0.60
ATOM  N2  Nhis  X   -0.65
ATOM  C17 CH1   X   0.51
ATOM  N6  NH2O  X   -0.12
ATOM  H17 Hpol  X   0.40
ATOM  H18 Hpol  X   0.40
ATOM  H6  Haro  X   0.06
ATOM  H7  Haro  X   0.06
ATOM  O5  OH    X   -0.56
ATOM  H14 Hapo  X   0.00
ATOM  O8  OH    X   -0.68
ATOM  H22 Hpol  X   0.40
ATOM  H12 Hapo  X   0.00
ATOM  O7  OH    X   -0.68
ATOM  H21 Hpol  X   0.40
ATOM  H11 Hapo  X   0.00
ATOM  H13 Hapo  X   0.00
ATOM  H15 Hapo  X   0.00
ATOM  H16 Hapo  X   0.00
ATOM  H19 Hpol  X   0.42
ATOM  O1  ONH2  X   -0.57
BOND_TYPE  C1   C2  4
BOND_TYPE  C1   C5  4
BOND_TYPE  C2   C7  4
BOND_TYPE  C3   C6  4
BOND_TYPE  C3   C7  4
BOND_TYPE  C4   C8  2
BOND_TYPE  C4   S1  1
BOND_TYPE  C5   C6  4
BOND_TYPE  C5   C13 1
BOND_TYPE  C6   O4  1
BOND_TYPE  C7   O6  1
BOND_TYPE  C8   C21 1
BOND_TYPE  C8   N1  1
BOND_TYPE  C9   C18 1
BOND_TYPE  C9   N1  2
BOND_TYPE  C9   S1  1
BOND_TYPE  C10  C13 2
BOND_TYPE  C10  C15 1
BOND_TYPE  C11  N2  2
BOND_TYPE  C11  N3  1
BOND_TYPE  C12  N4  1
```

```
BOND_TYPE  C12  N5  2
BOND_TYPE  C13  C27 1
BOND_TYPE  C14  C16 1
BOND_TYPE  C14  C17 1
BOND_TYPE  C14  N4  2
BOND_TYPE  C15  C19 2
BOND_TYPE  C15  O4  1
BOND_TYPE  C16  N3  2
BOND_TYPE  C16  N5  1
BOND_TYPE  C17  N2  1
BOND_TYPE  C17  N6  1
BOND_TYPE  C18  C20 2
BOND_TYPE  C19  C20 1
BOND_TYPE  C21  N7  4
BOND_TYPE  C21  O1  2
BOND_TYPE  C22  C23 1
BOND_TYPE  C22  C24 1
BOND_TYPE  C22  O7  1
BOND_TYPE  C23  C25 1
BOND_TYPE  C23  O8  1
BOND_TYPE  C24  C26 1
BOND_TYPE  C24  O5  1
BOND_TYPE  C25  N5  1
BOND_TYPE  C25  O5  1
BOND_TYPE  C26  O9  1
BOND_TYPE  C27  F1  1
BOND_TYPE  C27  F2  1
BOND_TYPE  C27  F3  1
BOND_TYPE  N7   S2  1
BOND_TYPE  O2   S2  2
BOND_TYPE  O3   S2  2
BOND_TYPE  O9   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C4   H4  1
BOND_TYPE  C10  H5  1
BOND_TYPE  C11  H6  1
BOND_TYPE  C12  H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C19  H9  1
BOND_TYPE  C20  H10 1
BOND_TYPE  C22  H11 1
BOND_TYPE  C23  H12 1
BOND_TYPE  C24  H13 1
BOND_TYPE  C25  H14 1
BOND_TYPE  C26  H15 1
BOND_TYPE  C26  H16 1
BOND_TYPE  N6   H17 1
BOND_TYPE  N6   H18 1
BOND_TYPE  N7   H19 1
BOND_TYPE  O6   H20 1
BOND_TYPE  O7   H21 1
BOND_TYPE  O8   H22 1
CHI 1  C2   C7   O6   H20
```

```
PROTON_CHI 1 SAMPLES 2 0 180 EXTRA 0
CHI 2  C24  C22  O7   H21
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 0
CHI 3  C22  C23  O8   H22
PROTON_CHI 3 SAMPLES 3 60 -60 180 EXTRA 0
CHI 4  N7   C21  C8   C4
CHI 5  S1   C9   C18  C20
CHI 6  C10  C13  C27  F1
CHI 7  C18  C20  C19  C15
CHI 8  O9   C26  C24  C22
CHI 9  C23  C25  N5   C12
CHI 10  S2   O9   C26  C24
CHI 11  C21  N7   S2   O2
CHI 12  N7   S2   O9   C26
NBR_ATOM  C21
NBR_RADIUS 16.869989
ICOOR_INTERNAL    C21    0.000000    0.000000    0.000000   C21   C8    C4
ICOOR_INTERNAL    C8     0.000000  180.000000    1.467142   C21   C8    C4
ICOOR_INTERNAL    C4     0.000000   53.603197    1.382530   C8    C21   C4
ICOOR_INTERNAL    S1  -179.253533   69.546811    1.714309   C4    C8    C21
ICOOR_INTERNAL    C9     0.329386   89.738119    1.729975   S1    C4    C8
ICOOR_INTERNAL    C18 -179.992720   54.987450    1.476047   C9    S1    C4
ICOOR_INTERNAL    C20  179.236740   57.142540    1.341649   C18   C9    S1
ICOOR_INTERNAL    C19  179.982036   59.264291    1.481055   C20   C18   C9
ICOOR_INTERNAL    C15 -179.996448   57.789060    1.343942   C19   C20   C18
ICOOR_INTERNAL    C10  179.999050   59.670344    1.483658   C15   C19   C20
ICOOR_INTERNAL    C13 -179.991608   59.833308    1.351148   C10   C15   C19
ICOOR_INTERNAL    C5    -0.004444   61.513320    1.499877   C13   C10   C15
ICOOR_INTERNAL    C1  -179.996621   57.435974    1.409032   C5    C13   C10
ICOOR_INTERNAL    C2   179.998465   59.522518    1.397663   C1    C5    C13
ICOOR_INTERNAL    C7    -0.004785   59.667532    1.400900   C2    C1    C5
ICOOR_INTERNAL    C3     0.002290   60.326517    1.397182   C7    C2    C1
ICOOR_INTERNAL    C6     0.000334   59.593488    1.398917   C3    C7    C2
ICOOR_INTERNAL    O4  -179.995549   60.901929    1.354343   C6    C3    C7
ICOOR_INTERNAL    H3  -179.998838   60.183084    1.083274   C3    C7    C6
ICOOR_INTERNAL    O6   179.996981   59.186210    1.348361   C7    C2    C3
ICOOR_INTERNAL    H20    0.003741   57.860590    0.967914   O6    C7    C2
ICOOR_INTERNAL    H2  -179.998714   60.540712    1.082888   C2    C1    C7
ICOOR_INTERNAL    H1   179.997594   58.496322    1.080365   C1    C5    C2
ICOOR_INTERNAL    C27  179.996201   59.614613    1.519435   C13   C10   C5
ICOOR_INTERNAL    F1   179.999197   67.754658    1.385061   C27   C13   C10
ICOOR_INTERNAL    F2  -120.033162   70.462510    1.382985   C27   C13   F1
ICOOR_INTERNAL    F3  -119.934302   70.463260    1.383016   C27   C13   F2
ICOOR_INTERNAL    H5   179.994447   61.221673    1.086269   C10   C15   C13
ICOOR_INTERNAL    H9   179.988904   61.465856    1.087703   C19   C20   C15
ICOOR_INTERNAL    H10 -179.987437   59.955724    1.085672   C20   C18   C19
ICOOR_INTERNAL    H8   179.984007   62.342041    1.087001   C18   C9    C20
ICOOR_INTERNAL    N1   179.286483   66.355734    1.317806   C9    S1    C18
ICOOR_INTERNAL    H4   179.166138   54.100665    1.081209   C4    C8    S1
ICOOR_INTERNAL    N7   179.999622   66.137129    1.468346   C21   C8    C4
ICOOR_INTERNAL    S2  -179.172745   58.297716    1.649429   N7    C21   C8
ICOOR_INTERNAL    O2   -83.826517   72.116620    1.437749   S2    N7    C21
ICOOR_INTERNAL    O3  -110.983663   68.091418    1.445053   S2    N7    O2
ICOOR_INTERNAL    O9  -129.798996   67.886550    1.508919   S2    N7    O3
ICOOR_INTERNAL    C26   47.974515   59.844760    1.411599   O9    S2    N7
```

```
ICOOR_INTERNAL    C24  157.659727   70.735303    1.511000   C26   O9    S2
ICOOR_INTERNAL    C22 -162.087031   70.934522    1.469223   C24   C26   O9
ICOOR_INTERNAL    C23 -144.252608   77.196882    1.465612   C22   C24   C26
ICOOR_INTERNAL    C25   31.365562   77.532435    1.475932   C23   C22   C24
ICOOR_INTERNAL    N5    91.986334   65.428356    1.445367   C25   C23   C22
ICOOR_INTERNAL    C12  -93.616101   52.734633    1.352203   N5    C25   C23
ICOOR_INTERNAL    N4  -179.129993   69.969290    1.324719   C12   N5    C25
ICOOR_INTERNAL    C14    0.432095   70.198470    1.328585   N4    C12   N5
ICOOR_INTERNAL    C16   -0.108979   73.546376    1.420752   C14   N4    C12
ICOOR_INTERNAL    N3   179.722517   60.629252    1.342146   C16   C14   N4
ICOOR_INTERNAL    C11    0.610914   59.624551    1.326052   N3    C16   C14
ICOOR_INTERNAL    N2    -0.850532   57.608301    1.331525   C11   N3    C16
ICOOR_INTERNAL    C17    0.549648   58.122539    1.342123   N2    C11   N3
ICOOR_INTERNAL    N6  -179.889809   60.506863    1.449372   C17   N2    C11
ICOOR_INTERNAL    H17  179.858300   60.000088    0.984533   N6    C17   N2
ICOOR_INTERNAL    H18 -179.994767   60.001461    0.984518   N6    C17   H17
ICOOR_INTERNAL    H6  -179.998716   61.190168    1.032048   C11   N3    N2
ICOOR_INTERNAL    H7   179.996347   55.015832    1.031901   C12   N5    N4
ICOOR_INTERNAL    O5  -122.232955   71.729727    1.413896   C25   C23   N5
ICOOR_INTERNAL    H14 -121.126627   73.236014    1.070090   C25   C23   O5
ICOOR_INTERNAL    O8  -116.509427   66.306380    1.374434   C23   C22   C25
ICOOR_INTERNAL    H22 -179.997499   70.525025    0.969944   O8    C23   C22
ICOOR_INTERNAL    H12 -118.788948   69.018463    1.070044   C23   C22   O8
ICOOR_INTERNAL    O7  -122.096059   69.952681    1.380155   C22   C24   C23
ICOOR_INTERNAL    H21  -64.965720   70.523320    0.970083   O7    C22   C24
ICOOR_INTERNAL    H11 -117.021116   64.832231    1.070051   C22   C24   O7
ICOOR_INTERNAL    H13  119.997304   70.465137    1.070087   C24   C26   C22
ICOOR_INTERNAL    H15 -119.965789   70.478627    1.070027   C26   O9    C24
ICOOR_INTERNAL    H16 -119.933086   70.407815    1.069992   C26   O9    H15
ICOOR_INTERNAL    H19  179.999381   60.851423    0.984579   N7    C21   S2
ICOOR_INTERNAL    O1  -177.781693   57.290531    1.227179   C21   C8    N7
PDB_ROTAMERS LCBb_conformers.pdb
```

The CouLuc-3-OH .params file for the E isomer is below.

```
NAME LCB
IO_STRING LCB Z
TYPE LIGAND
AA UNK
ATOM  C21 CNH2  X   0.89
ATOM  C8  aroC  X   0.14
ATOM  C4  aroC  X   -0.11
ATOM  S1  S     X   -0.08
ATOM  C9  aroC  X   0.33
ATOM  C18 aroC  X   -0.10
ATOM  C20 aroC  X   -0.15
ATOM  C19 aroC  X   -0.15
ATOM  C15 aroC  X   0.08
ATOM  C10 aroC  X   -0.15
ATOM  C13 aroC  X   -0.17
ATOM  C5  aroC  X   0.03
ATOM  C1  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C7  aroC  X   0.08
ATOM  C3  aroC  X   -0.15
```

```
ATOM  C6  aroC  X   0.08
ATOM  O4  Oaro  X   -0.16
ATOM  H3  Haro  X   0.15
ATOM  O6  OH    X   -0.53
ATOM  H20 Hpol  X   0.45
ATOM  H2  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  C27 CH1   X   1.16
ATOM  F1  F     X   -0.34
ATOM  F2  F     X   -0.34
ATOM  F3  F     X   -0.34
ATOM  H5  Haro  X   0.15
ATOM  H9  Haro  X   0.15
ATOM  H10 Haro  X   0.15
ATOM  H8  Haro  X   0.15
ATOM  N1  Nhis  X   -0.57
ATOM  H4  Haro  X   0.15
ATOM  N7  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O2  OOC   X   -0.65
ATOM  O3  OOC   X   -0.65
ATOM  O9  OH    X   -0.46
ATOM  C26 CH2   X   0.28
ATOM  C24 CH1   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C23 CH1   X   0.28
ATOM  C25 CH1   X   0.63
ATOM  N5  Npro  X   -0.08
ATOM  C12 aroC  X   0.55
ATOM  N4  Nhis  X   -0.66
ATOM  C14 aroC  X   0.46
ATOM  C16 aroC  X   0.78
ATOM  N3  Nhis  X   -0.66
ATOM  C11 aroC  X   0.60
ATOM  N2  Nhis  X   -0.65
ATOM  C17 CH1   X   0.51
ATOM  N6  NH2O  X   -0.12
ATOM  H17 Hpol  X   0.40
ATOM  H18 Hpol  X   0.40
ATOM  H6  Haro  X   0.06
ATOM  H7  Haro  X   0.06
ATOM  O5  OH    X   -0.56
ATOM  H14 Hapo  X   0.00
ATOM  O8  OH    X   -0.68
ATOM  H22 Hpol  X   0.40
ATOM  H12 Hapo  X   0.00
ATOM  O7  OH    X   -0.68
ATOM  H21 Hpol  X   0.40
ATOM  H11 Hapo  X   0.00
ATOM  H13 Hapo  X   0.00
ATOM  H15 Hapo  X   0.00
ATOM  H16 Hapo  X   0.00
ATOM  H19 Hpol  X   0.42
ATOM  O1  ONH2  X   -0.57
BOND_TYPE  C1   C2  4
```

```
BOND_TYPE  C1   C5  4
BOND_TYPE  C2   C7  4
BOND_TYPE  C3   C6  4
BOND_TYPE  C3   C7  4
BOND_TYPE  C4   C8  2
BOND_TYPE  C4   S1  1
BOND_TYPE  C5   C6  4
BOND_TYPE  C5   C13 1
BOND_TYPE  C6   O4  1
BOND_TYPE  C7   O6  1
BOND_TYPE  C8   C21 1
BOND_TYPE  C8   N1  1
BOND_TYPE  C9   C18 1
BOND_TYPE  C9   N1  2
BOND_TYPE  C9   S1  1
BOND_TYPE  C10  C13 2
BOND_TYPE  C10  C15 1
BOND_TYPE  C11  N2  2
BOND_TYPE  C11  N3  1
BOND_TYPE  C12  N4  1
BOND_TYPE  C12  N5  2
BOND_TYPE  C13  C27 1
BOND_TYPE  C14  C16 1
BOND_TYPE  C14  C17 1
BOND_TYPE  C14  N4  2
BOND_TYPE  C15  C19 2
BOND_TYPE  C15  O4  1
BOND_TYPE  C16  N3  2
BOND_TYPE  C16  N5  1
BOND_TYPE  C17  N2  1
BOND_TYPE  C17  N6  1
BOND_TYPE  C18  C20 2
BOND_TYPE  C19  C20 1
BOND_TYPE  C21  N7  4
BOND_TYPE  C21  O1  2
BOND_TYPE  C22  C23 1
BOND_TYPE  C22  C24 1
BOND_TYPE  C22  O7  1
BOND_TYPE  C23  C25 1
BOND_TYPE  C23  O8  1
BOND_TYPE  C24  C26 1
BOND_TYPE  C24  O5  1
BOND_TYPE  C25  N5  1
BOND_TYPE  C25  O5  1
BOND_TYPE  C26  O9  1
BOND_TYPE  C27  F1  1
BOND_TYPE  C27  F2  1
BOND_TYPE  C27  F3  1
BOND_TYPE  N7   S2  1
BOND_TYPE  O2   S2  2
BOND_TYPE  O3   S2  2
BOND_TYPE  O9   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
```

```
BOND_TYPE  C4   H4  1
BOND_TYPE  C10  H5  1
BOND_TYPE  C11  H6  1
BOND_TYPE  C12  H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C19  H9  1
BOND_TYPE  C20  H10 1
BOND_TYPE  C22  H11 1
BOND_TYPE  C23  H12 1
BOND_TYPE  C24  H13 1
BOND_TYPE  C25  H14 1
BOND_TYPE  C26  H15 1
BOND_TYPE  C26  H16 1
BOND_TYPE  N6   H17 1
BOND_TYPE  N6   H18 1
BOND_TYPE  N7   H19 1
BOND_TYPE  O6   H20 1
BOND_TYPE  O7   H21 1
BOND_TYPE  O8   H22 1
CHI 1  C2   C7   O6   H20
PROTON_CHI 1 SAMPLES 2 0 180 EXTRA 0
CHI 2  C24  C22  O7   H21
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 0
CHI 3  C22  C23  O8   H22
PROTON_CHI 3 SAMPLES 3 60 -60 180 EXTRA 0
CHI 4  N7   C21  C8   C4
CHI 5  S1   C9   C18  C20
CHI 6  C10  C13  C27  F1
CHI 7  C18  C20  C19  C15
CHI 8  O9   C26  C24  C22
CHI 9  C23  C25  N5   C12
CHI 10  S2   O9   C26  C24
CHI 11  C21  N7   S2   O2
CHI 12  N7   S2   O9   C26
NBR_ATOM  C21
NBR_RADIUS 16.869989
ICOOR_INTERNAL    C21    0.000000    0.000000    0.000000   C21   C8    C4
ICOOR_INTERNAL    C8     0.000000  180.000000    1.467142   C21   C8    C4
ICOOR_INTERNAL    C4     0.000000   53.603197    1.382530   C8    C21   C4
ICOOR_INTERNAL    S1  -179.253533   69.546811    1.714309   C4    C8    C21
ICOOR_INTERNAL    C9     0.329386   89.738119    1.729975   S1    C4    C8
ICOOR_INTERNAL    C18  174.789967   60.958622    1.472921   C9    S1    C4
ICOOR_INTERNAL    C20 -174.817607   52.374999    1.341456   C18   C9    S1
ICOOR_INTERNAL    C19  176.669795   63.190030    1.489780   C20   C18   C9
ICOOR_INTERNAL    C15  179.999613   52.099225    1.343116   C19   C20   C18
ICOOR_INTERNAL    C10   -3.940119   55.615188    1.487039   C15   C19   C20
ICOOR_INTERNAL    C13 -178.767537   60.324351    1.356134   C10   C15   C19
ICOOR_INTERNAL    C5    -1.908797   61.119747    1.501774   C13   C10   C15
ICOOR_INTERNAL    C1  -177.946289   57.079642    1.409867   C5    C13   C10
ICOOR_INTERNAL    C2   179.974256   59.489131    1.399093   C1    C5    C13
ICOOR_INTERNAL    C7     0.123786   59.683863    1.400785   C2    C1    C5
ICOOR_INTERNAL    C3    -0.006688   60.396556    1.396000   C7    C2    C1
ICOOR_INTERNAL    C6    -0.112866   59.618090    1.399076   C3    C7    C2
ICOOR_INTERNAL    O4  -179.840928   60.986484    1.350060   C6    C3    C7
ICOOR_INTERNAL    H3  -179.857894   60.167913    1.083461   C3    C7    C6
```

```
ICOOR_INTERNAL    O6  -179.963095   59.088533    1.348698   C7    C2    C3
ICOOR_INTERNAL    H20    0.008392   57.812798    0.967746   O6    C7    C2
ICOOR_INTERNAL    H2  -179.987008   60.467903    1.082729   C2    C1    C7
ICOOR_INTERNAL    H1  -179.811248   58.444172    1.080036   C1    C5    C2
ICOOR_INTERNAL    C27  179.219739   59.314271    1.518084   C13   C10   C5
ICOOR_INTERNAL    F1  -179.212895   67.765697    1.382974   C27   C13   C10
ICOOR_INTERNAL    F2  -119.216427   70.652560    1.383140   C27   C13   F1
ICOOR_INTERNAL    F3  -119.858809   70.297200    1.382780   C27   C13   F2
ICOOR_INTERNAL    H5   178.116472   60.466234    1.083816   C10   C15   C13
ICOOR_INTERNAL    H9   179.518166   63.514338    1.088961   C19   C20   C15
ICOOR_INTERNAL    H10 -179.436665   59.109031    1.084693   C20   C18   C19
ICOOR_INTERNAL    H8   179.998924   65.772904    1.087972   C18   C9    C20
ICOOR_INTERNAL    N1  -175.496204   66.355734    1.317806   C9    S1    C18
ICOOR_INTERNAL    H4   178.032084   50.570774    1.080736   C4    C8    S1
ICOOR_INTERNAL    N7   179.999622   66.137129    1.468346   C21   C8    C4
ICOOR_INTERNAL    S2  -179.172745   58.297716    1.649429   N7    C21   C8
ICOOR_INTERNAL    O2   -83.826517   72.116620    1.437749   S2    N7    C21
ICOOR_INTERNAL    O3  -110.983663   68.091418    1.445053   S2    N7    O2
ICOOR_INTERNAL    O9  -129.798996   67.886550    1.508919   S2    N7    O3
ICOOR_INTERNAL    C26   47.974515   59.844760    1.411599   O9    S2    N7
ICOOR_INTERNAL    C24  157.659727   70.735303    1.511000   C26   O9    S2
ICOOR_INTERNAL    C22 -162.087031   70.934522    1.469223   C24   C26   O9
ICOOR_INTERNAL    C23 -144.252608   77.196882    1.465612   C22   C24   C26
ICOOR_INTERNAL    C25   31.365562   77.532435    1.475932   C23   C22   C24
ICOOR_INTERNAL    N5    91.986334   65.428356    1.445367   C25   C23   C22
ICOOR_INTERNAL    C12  -93.616101   52.734633    1.352203   N5    C25   C23
ICOOR_INTERNAL    N4  -179.129993   69.969290    1.324719   C12   N5    C25
ICOOR_INTERNAL    C14    0.432095   70.198470    1.328585   N4    C12   N5
ICOOR_INTERNAL    C16   -0.108979   73.546376    1.420752   C14   N4    C12
ICOOR_INTERNAL    N3   179.722517   60.629252    1.342146   C16   C14   N4
ICOOR_INTERNAL    C11    0.610914   59.624551    1.326052   N3    C16   C14
ICOOR_INTERNAL    N2    -0.850532   57.608301    1.331525   C11   N3    C16
ICOOR_INTERNAL    C17    0.549648   58.122539    1.342123   N2    C11   N3
ICOOR_INTERNAL    N6  -179.889809   60.506863    1.449372   C17   N2    C11
ICOOR_INTERNAL    H17  179.858300   60.000088    0.984533   N6    C17   N2
ICOOR_INTERNAL    H18 -179.994767   60.001461    0.984518   N6    C17   H17
ICOOR_INTERNAL    H6  -179.998716   61.190168    1.032048   C11   N3    N2
ICOOR_INTERNAL    H7   179.996347   55.015832    1.031901   C12   N5    N4
ICOOR_INTERNAL    O5  -122.232955   71.729727    1.413896   C25   C23   N5
ICOOR_INTERNAL    H14 -121.126627   73.236014    1.070090   C25   C23   O5
ICOOR_INTERNAL    O8  -116.509427   66.306380    1.374434   C23   C22   C25
ICOOR_INTERNAL    H22 -179.997499   70.525025    0.969944   O8    C23   C22
ICOOR_INTERNAL    H12 -118.788948   69.018463    1.070044   C23   C22   O8
ICOOR_INTERNAL    O7  -122.096059   69.952681    1.380155   C22   C24   C23
ICOOR_INTERNAL    H21  -64.965720   70.523320    0.970083   O7    C22   C24
ICOOR_INTERNAL    H11 -117.021116   64.832231    1.070051   C22   C24   O7
ICOOR_INTERNAL    H13  119.997304   70.465137    1.070087   C24   C26   C22
ICOOR_INTERNAL    H15 -119.965789   70.478627    1.070027   C26   O9    C24
ICOOR_INTERNAL    H16 -119.933086   70.407815    1.069992   C26   O9    H15
ICOOR_INTERNAL    H19  179.999381   60.851423    0.984579   N7    C21   S2
ICOOR_INTERNAL    O1  -177.781693   57.290531    1.227179   C21   C8    N7
PDB_ROTAMERS LCBa_conformers.pdb
```

The contents of the CouLuc-3-NH_2.params file for the E isomer is below.

```
NAME LCC
IO_STRING LCC Z
TYPE LIGAND
AA UNK
ATOM  C21 CNH2  X   0.89
ATOM  C8  aroC  X   0.14
ATOM  C4  aroC  X   -0.11
ATOM  S1  S     X   -0.08
ATOM  C9  aroC  X   0.33
ATOM  C18 aroC  X   -0.10
ATOM  C20 aroC  X   -0.15
ATOM  C19 aroC  X   -0.15
ATOM  C15 aroC  X   0.08
ATOM  C10 aroC  X   -0.15
ATOM  C13 aroC  X   -0.17
ATOM  C5  aroC  X   0.03
ATOM  C1  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C6  aroC  X   0.10
ATOM  C3  aroC  X   -0.15
ATOM  C7  aroC  X   0.08
ATOM  O4  Oaro  X   -0.16
ATOM  H3  Haro  X   0.15
ATOM  N6  NH2O  X   -0.90
ATOM  H17 Hpol  X   0.40
ATOM  H18 Hpol  X   0.40
ATOM  H2  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  C27 CH1   X   1.16
ATOM  F1  F     X   -0.34
ATOM  F2  F     X   -0.34
ATOM  F3  F     X   -0.34
ATOM  H5  Haro  X   0.15
ATOM  H9  Haro  X   0.15
ATOM  H10 Haro  X   0.15
ATOM  H8  Haro  X   0.15
ATOM  N1  Nhis  X   -0.57
ATOM  H4  Haro  X   0.15
ATOM  N8  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O2  OOC   X   -0.65
ATOM  O3  OOC   X   -0.65
ATOM  O8  OH    X   -0.46
ATOM  C26 CH2   X   0.28
ATOM  C24 CH1   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C23 CH1   X   0.28
ATOM  C25 CH1   X   0.63
ATOM  N5  Npro  X   -0.08
ATOM  C12 aroC  X   0.55
ATOM  N4  Nhis  X   -0.66
ATOM  C14 aroC  X   0.46
ATOM  C16 aroC  X   0.78
ATOM  N3  Nhis  X   -0.66
```

```
ATOM  C11 aroC  X   0.60
ATOM  N2  Nhis  X   -0.65
ATOM  C17 CH1   X   0.51
ATOM  N7  NH2O  X   -0.12
ATOM  H19 Hpol  X   0.40
ATOM  H20 Hpol  X   0.40
ATOM  H6  Haro  X   0.06
ATOM  H7  Haro  X   0.06
ATOM  O5  OH    X   -0.56
ATOM  H14 Hapo  X   0.00
ATOM  O7  OH    X   -0.68
ATOM  H23 Hpol  X   0.40
ATOM  H12 Hapo  X   0.00
ATOM  O6  OH    X   -0.68
ATOM  H22 Hpol  X   0.40
ATOM  H11 Hapo  X   0.00
ATOM  H13 Hapo  X   0.00
ATOM  H15 Hapo  X   0.00
ATOM  H16 Hapo  X   0.00
ATOM  H21 Hpol  X   0.42
ATOM  O1  ONH2  X   -0.57
BOND_TYPE  C1   C2  4
BOND_TYPE  C1   C5  4
BOND_TYPE  C2   C6  4
BOND_TYPE  C3   C6  4
BOND_TYPE  C3   C7  4
BOND_TYPE  C4   C8  2
BOND_TYPE  C4   S1  1
BOND_TYPE  C5   C7  4
BOND_TYPE  C5   C13 1
BOND_TYPE  C6   N6  1
BOND_TYPE  C7   O4  1
BOND_TYPE  C8   C21 1
BOND_TYPE  C8   N1  1
BOND_TYPE  C9   C18 1
BOND_TYPE  C9   N1  2
BOND_TYPE  C9   S1  1
BOND_TYPE  C10  C13 2
BOND_TYPE  C10  C15 1
BOND_TYPE  C11  N2  2
BOND_TYPE  C11  N3  1
BOND_TYPE  C12  N4  1
BOND_TYPE  C12  N5  2
BOND_TYPE  C13  C27 1
BOND_TYPE  C14  C16 1
BOND_TYPE  C14  C17 1
BOND_TYPE  C14  N4  2
BOND_TYPE  C15  C19 2
BOND_TYPE  C15  O4  1
BOND_TYPE  C16  N3  2
BOND_TYPE  C16  N5  1
BOND_TYPE  C17  N2  1
BOND_TYPE  C17  N7  1
BOND_TYPE  C18  C20 2
BOND_TYPE  C19  C20 1
```

```
BOND_TYPE  C21  N8  4
BOND_TYPE  C21  O1  2
BOND_TYPE  C22  C23 1
BOND_TYPE  C22  C24 1
BOND_TYPE  C22  O6  1
BOND_TYPE  C23  C25 1
BOND_TYPE  C23  O7  1
BOND_TYPE  C24  C26 1
BOND_TYPE  C24  O5  1
BOND_TYPE  C25  N5  1
BOND_TYPE  C25  O5  1
BOND_TYPE  C26  O8  1
BOND_TYPE  C27  F1  1
BOND_TYPE  C27  F2  1
BOND_TYPE  C27  F3  1
BOND_TYPE  N8   S2  1
BOND_TYPE  O2   S2  2
BOND_TYPE  O3   S2  2
BOND_TYPE  O8   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C4   H4  1
BOND_TYPE  C10  H5  1
BOND_TYPE  C11  H6  1
BOND_TYPE  C12  H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C19  H9  1
BOND_TYPE  C20  H10 1
BOND_TYPE  C22  H11 1
BOND_TYPE  C23  H12 1
BOND_TYPE  C24  H13 1
BOND_TYPE  C25  H14 1
BOND_TYPE  C26  H15 1
BOND_TYPE  C26  H16 1
BOND_TYPE  N6   H17 1
BOND_TYPE  N6   H18 1
BOND_TYPE  N7   H19 1
BOND_TYPE  N7   H20 1
BOND_TYPE  N8   H21 1
BOND_TYPE  O6   H22 1
BOND_TYPE  O7   H23 1
CHI 1  C24  C22  O6   H22
PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 0
CHI 2  C22  C23  O7   H23
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 0
CHI 3  N8   C21  C8   C4
CHI 4  S1   C9   C18  C20
CHI 5  C10  C13  C27  F1
CHI 6  C18  C20  C19  C15
CHI 7  O8   C26  C24  C22
CHI 8  C23  C25  N5   C12
CHI 9  S2   O8   C26  C24
CHI 10  C21  N8   S2   O2
CHI 11  N8   S2   O8   C26
```

```
NBR_ATOM  C21
NBR_RADIUS 16.870172
ICOOR_INTERNAL    C21    0.000000    0.000000    0.000000   C21   C8
C4
ICOOR_INTERNAL    C8     0.000000  180.000000    1.467099   C21   C8
C4
ICOOR_INTERNAL    C4     0.000001   53.613584    1.382650   C8    C21
C4
ICOOR_INTERNAL    S1  -
179.277244   69.545521    1.714033   C4    C8    C21
ICOOR_INTERNAL    C9     0.366000   89.733106    1.730169   S1    C4
C8
ICOOR_INTERNAL    C18  178.553405   53.909816    1.470946   C9    S1
C4
ICOOR_INTERNAL    C20 -
179.324030   55.671334    1.340708   C18   C9    S1
ICOOR_INTERNAL    C19 -
178.869499   61.124213    1.479088   C20   C18   C9
ICOOR_INTERNAL    C15  179.996690   56.167418    1.341760   C19   C20
C18
ICOOR_INTERNAL    C10 -
179.328007   60.815784    1.477883   C15   C19   C20
ICOOR_INTERNAL    C13  179.018646   58.951715    1.346924   C10   C15
C19
ICOOR_INTERNAL    C5     0.083660   62.211270    1.512953   C13   C10
C15
ICOOR_INTERNAL    C1  -
179.720574   56.106976    1.406783   C5    C13   C10
ICOOR_INTERNAL    C2  -
179.792112   57.923251    1.397442   C1    C5    C13
ICOOR_INTERNAL    C6    -
0.121670   60.422992    1.400044   C2    C1    C5
ICOOR_INTERNAL    C3    -
0.014294   60.447619    1.402716   C6    C2    C1
ICOOR_INTERNAL    C7     0.064356   59.281962    1.401587   C3    C6
C2
ICOOR_INTERNAL    O4  -
179.893959   61.173171    1.357107   C7    C3    C6
ICOOR_INTERNAL    H3  -
179.996790   60.096543    1.083290   C3    C6    C7
ICOOR_INTERNAL    N6  -
179.957136   59.786317    1.416636   C6    C2    C3
ICOOR_INTERNAL    H17   -
0.010267   59.286668    1.030553   N6    C6    C2
ICOOR_INTERNAL    H18 -
179.981802   59.285011    1.030602   N6    C6    H17
ICOOR_INTERNAL    H2  -
179.960013   60.099881    1.083143   C2    C1    C6
ICOOR_INTERNAL    H1   179.883915   56.831332    1.070336   C1    C5
C2
ICOOR_INTERNAL    C27 -
179.974949   64.668326    1.528344   C13   C10   C5
ICOOR_INTERNAL    F1   179.970880   64.860846    1.374701   C27   C13
C10
```

```
ICOOR_INTERNAL    F2   120.803880   70.973223    1.384120   C27   C13
F1
ICOOR_INTERNAL    F3   118.347981   70.969959    1.384140   C27   C13
F2
ICOOR_INTERNAL    H5  -
179.791393   61.100806    1.088681   C10   C15   C13
ICOOR_INTERNAL    H9  -
179.288766   62.277338    1.088530   C19   C20   C15
ICOOR_INTERNAL    H10  179.084778   59.182765    1.086090   C20   C18
C19
ICOOR_INTERNAL    H8  -
178.987338   63.499105    1.087867   C18   C9    C20
ICOOR_INTERNAL    N1  -
179.366714   66.372571    1.317944   C9    S1    C18
ICOOR_INTERNAL    H4   178.052410   50.569171    1.080507   C4    C8
S1
ICOOR_INTERNAL    N8   179.999853   66.190937    1.468440   C21   C8
C4
ICOOR_INTERNAL    S2  -
179.026765   58.301303    1.649468   N8    C21   C8
ICOOR_INTERNAL    O2   -
83.829697   72.117552    1.437687   S2    N8    C21
ICOOR_INTERNAL    O3  -
110.978610   68.096527    1.445149   S2    N8    O2
ICOOR_INTERNAL    O8  -
129.796645   67.882755    1.508902   S2    N8    O3
ICOOR_INTERNAL    C26   47.972090   59.846860    1.411589   O8    S2
N8
ICOOR_INTERNAL    C24  157.664803   70.737827    1.511020   C26   O8
S2
ICOOR_INTERNAL    C22 -
162.089472   70.934687    1.469290   C24   C26   O8
ICOOR_INTERNAL    C23 -
144.250635   77.197183    1.465518   C22   C24   C26
ICOOR_INTERNAL    C25   31.359911   77.525336    1.475945   C23   C22
C24
ICOOR_INTERNAL    N5    91.982589   65.434427    1.445330   C25   C23
C22
ICOOR_INTERNAL    C12  -
93.612286   52.726401    1.352130   N5    C25   C23
ICOOR_INTERNAL    N4  -
179.125529   69.963424    1.324697   C12   N5    C25
ICOOR_INTERNAL    C14    0.434724   70.201937    1.328569   N4    C12
N5
ICOOR_INTERNAL    C16   -
0.108966   73.543980    1.420687   C14   N4    C12
ICOOR_INTERNAL    N3   179.727076   60.631876    1.342184   C16   C14
N4
ICOOR_INTERNAL    C11    0.605797   59.617650    1.325985   N3    C16
C14
ICOOR_INTERNAL    N2    -
0.843694   57.615869    1.331578   C11   N3    C16
ICOOR_INTERNAL    C17    0.543846   58.116997    1.342047   N2    C11
N3
```

```
ICOOR_INTERNAL    N7  -
179.885925   60.505121    1.449333   C17   N2    C11
ICOOR_INTERNAL    H19  179.857903   59.996676    0.984436   N7    C17
N2
ICOOR_INTERNAL    H20 -
179.996362   60.003205    0.984528   N7    C17   H19
ICOOR_INTERNAL    H6   179.991347   61.186872    1.031972   C11   N3
N2
ICOOR_INTERNAL    H7   179.994533   55.014941    1.032043   C12   N5
N4
ICOOR_INTERNAL    O5  -
122.220593   71.733603    1.413878   C25   C23   N5
ICOOR_INTERNAL    H14 -
121.131406   73.232551    1.069922   C25   C23   O5
ICOOR_INTERNAL    O7  -
116.507140   66.309831    1.374491   C23   C22   C25
ICOOR_INTERNAL    H23  179.997609   70.527483    0.969954   O7    C23
C22
ICOOR_INTERNAL    H12 -
120.201693   66.638819    1.112365   C23   C22   O7
ICOOR_INTERNAL    O6  -
122.102731   69.950997    1.380152   C22   C24   C23
ICOOR_INTERNAL    H22  -
64.962631   70.528340    0.970025   O6    C22   C24
ICOOR_INTERNAL    H11 -
117.021009   64.837781    1.070175   C22   C24   O6
ICOOR_INTERNAL    H13  119.990690   70.467539    1.070000   C24   C26
C22
ICOOR_INTERNAL    H15 -
119.973622   70.474860    1.069935   C26   O8    C24
ICOOR_INTERNAL    H16 -
119.929165   70.409662    1.070081   C26   O8    H15
ICOOR_INTERNAL    H21  179.995484   60.852845    0.984556   N8    C21
S2
ICOOR_INTERNAL    O1  -
177.629925   57.236261    1.227139   C21   C8    N8
PDB_ROTAMERS LCCb_conformers.pdb
```

The contents of the CouLuc-3-NH_2 .params file for the Z isomer is below.

```
NAME LCC
IO_STRING LCC Z
TYPE LIGAND
AA UNK
ATOM  C21 CNH2  X   0.89
ATOM  C8  aroC  X   0.14
ATOM  C4  aroC  X   -0.11
ATOM  S1  S     X   -0.08
ATOM  C9  aroC  X   0.33
ATOM  C18 aroC  X   -0.10
ATOM  C20 aroC  X   -0.15
ATOM  C19 aroC  X   -0.15
ATOM  C15 aroC  X   0.08
ATOM  C10 aroC  X   -0.15
ATOM  C13 aroC  X   -0.17
```

```
ATOM  C5  aroC  X   0.03
ATOM  C1  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C6  aroC  X   0.10
ATOM  C3  aroC  X   -0.15
ATOM  C7  aroC  X   0.08
ATOM  O4  Oaro  X   -0.16
ATOM  H3  Haro  X   0.15
ATOM  N6  NH2O  X   -0.90
ATOM  H17 Hpol  X   0.40
ATOM  H18 Hpol  X   0.40
ATOM  H2  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  C27 CH1   X   1.16
ATOM  F1  F     X   -0.34
ATOM  F2  F     X   -0.34
ATOM  F3  F     X   -0.34
ATOM  H5  Haro  X   0.15
ATOM  H9  Haro  X   0.15
ATOM  H10 Haro  X   0.15
ATOM  H8  Haro  X   0.15
ATOM  N1  Nhis  X   -0.57
ATOM  H4  Haro  X   0.15
ATOM  N8  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O2  OOC   X   -0.65
ATOM  O3  OOC   X   -0.65
ATOM  O8  OH    X   -0.46
ATOM  C26 CH2   X   0.28
ATOM  C24 CH1   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C23 CH1   X   0.28
ATOM  C25 CH1   X   0.63
ATOM  N5  Npro  X   -0.08
ATOM  C12 aroC  X   0.55
ATOM  N4  Nhis  X   -0.66
ATOM  C14 aroC  X   0.46
ATOM  C16 aroC  X   0.78
ATOM  N3  Nhis  X   -0.66
ATOM  C11 aroC  X   0.60
ATOM  N2  Nhis  X   -0.65
ATOM  C17 CH1   X   0.51
ATOM  N7  NH2O  X   -0.12
ATOM  H19 Hpol  X   0.40
ATOM  H20 Hpol  X   0.40
ATOM  H6  Haro  X   0.06
ATOM  H7  Haro  X   0.06
ATOM  O5  OH    X   -0.56
ATOM  H14 Hapo  X   0.00
ATOM  O7  OH    X   -0.68
ATOM  H23 Hpol  X   0.40
ATOM  H12 Hapo  X   0.00
ATOM  O6  OH    X   -0.68
ATOM  H22 Hpol  X   0.40
ATOM  H11 Hapo  X   0.00
```

```
ATOM  H13 Hapo  X   0.00
ATOM  H15 Hapo  X   0.00
ATOM  H16 Hapo  X   0.00
ATOM  H21 Hpol  X   0.42
ATOM  O1  ONH2  X   -0.57
BOND_TYPE  C1   C2  4
BOND_TYPE  C1   C5  4
BOND_TYPE  C2   C6  4
BOND_TYPE  C3   C6  4
BOND_TYPE  C3   C7  4
BOND_TYPE  C4   C8  2
BOND_TYPE  C4   S1  1
BOND_TYPE  C5   C7  4
BOND_TYPE  C5   C13 1
BOND_TYPE  C6   N6  1
BOND_TYPE  C7   O4  1
BOND_TYPE  C8   C21 1
BOND_TYPE  C8   N1  1
BOND_TYPE  C9   C18 1
BOND_TYPE  C9   N1  2
BOND_TYPE  C9   S1  1
BOND_TYPE  C10  C13 2
BOND_TYPE  C10  C15 1
BOND_TYPE  C11  N2  2
BOND_TYPE  C11  N3  1
BOND_TYPE  C12  N4  1
BOND_TYPE  C12  N5  2
BOND_TYPE  C13  C27 1
BOND_TYPE  C14  C16 1
BOND_TYPE  C14  C17 1
BOND_TYPE  C14  N4  2
BOND_TYPE  C15  C19 2
BOND_TYPE  C15  O4  1
BOND_TYPE  C16  N3  2
BOND_TYPE  C16  N5  1
BOND_TYPE  C17  N2  1
BOND_TYPE  C17  N7  1
BOND_TYPE  C18  C20 2
BOND_TYPE  C19  C20 1
BOND_TYPE  C21  N8  4
BOND_TYPE  C21  O1  2
BOND_TYPE  C22  C23 1
BOND_TYPE  C22  C24 1
BOND_TYPE  C22  O6  1
BOND_TYPE  C23  C25 1
BOND_TYPE  C23  O7  1
BOND_TYPE  C24  C26 1
BOND_TYPE  C24  O5  1
BOND_TYPE  C25  N5  1
BOND_TYPE  C25  O5  1
BOND_TYPE  C26  O8  1
BOND_TYPE  C27  F1  1
BOND_TYPE  C27  F2  1
BOND_TYPE  C27  F3  1
BOND_TYPE  N8   S2  1
```

```
BOND_TYPE  O2   S2  2
BOND_TYPE  O3   S2  2
BOND_TYPE  O8   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C4   H4  1
BOND_TYPE  C10  H5  1
BOND_TYPE  C11  H6  1
BOND_TYPE  C12  H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C19  H9  1
BOND_TYPE  C20  H10 1
BOND_TYPE  C22  H11 1
BOND_TYPE  C23  H12 1
BOND_TYPE  C24  H13 1
BOND_TYPE  C25  H14 1
BOND_TYPE  C26  H15 1
BOND_TYPE  C26  H16 1
BOND_TYPE  N6   H17 1
BOND_TYPE  N6   H18 1
BOND_TYPE  N7   H19 1
BOND_TYPE  N7   H20 1
BOND_TYPE  N8   H21 1
BOND_TYPE  O6   H22 1
BOND_TYPE  O7   H23 1
CHI 1  C24  C22  O6   H22
PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 0
CHI 2  C22  C23  O7   H23
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 0
CHI 3  N8   C21  C8   C4
CHI 4  S1   C9   C18  C20
CHI 5  C10  C13  C27  F1
CHI 6  C18  C20  C19  C15
CHI 7  O8   C26  C24  C22
CHI 8  C23  C25  N5   C12
CHI 9  S2   O8   C26  C24
CHI 10  C21  N8   S2   O2
CHI 11  N8   S2   O8   C26
NBR_ATOM  C21
NBR_RADIUS 16.870131
ICOOR_INTERNAL    C21    0.000000    0.000000    0.000000   C21   C8    C4
ICOOR_INTERNAL    C8     0.000000  179.999999    1.467129   C21   C8    C4
ICOOR_INTERNAL    C4     0.000001   53.617032    1.382669   C8    C21   C4
ICOOR_INTERNAL    S1  -179.281387   69.551912    1.714111   C4    C8    C21
ICOOR_INTERNAL    C9     0.368538   89.731913    1.730053   S1    C4    C8
ICOOR_INTERNAL    C18 -179.983428   55.061880    1.476720   C9    S1    C4
ICOOR_INTERNAL    C20  179.108433   57.299028    1.341098   C18   C9    S1
ICOOR_INTERNAL    C19 -179.997376   59.599319    1.483349   C20   C18   C9
ICOOR_INTERNAL    C15  179.998708   54.652274    1.347872   C19   C20   C18
ICOOR_INTERNAL    C10   -0.025247   56.152013    1.492261   C15   C19   C20
ICOOR_INTERNAL    C13 -179.987967   59.792909    1.351295   C10   C15   C19
ICOOR_INTERNAL    C5     0.003565   61.420008    1.497790   C13   C10   C15
ICOOR_INTERNAL    C1   179.993562   57.471192    1.407653   C5    C13   C10
ICOOR_INTERNAL    C2   179.998015   59.547970    1.396492   C1    C5    C13
```

```
ICOOR_INTERNAL    C6    -0.001108   59.464747    1.399978   C2    C1    C5
ICOOR_INTERNAL    C3     0.005501   60.566274    1.399914   C6    C2    C1
ICOOR_INTERNAL    C7     0.002166   59.609795    1.398888   C3    C6    C2
ICOOR_INTERNAL    O4   179.998709   60.952031    1.353058   C7    C3    C6
ICOOR_INTERNAL    H3   179.989981   59.941927    1.083245   C3    C6    C7
ICOOR_INTERNAL    N6   179.998939   59.722662    1.416621   C6    C2    C3
ICOOR_INTERNAL    H17    0.000439   59.291948    1.030503   N6    C6    C2
ICOOR_INTERNAL    H18  179.996618   59.286299    1.030608   N6    C6    H17
ICOOR_INTERNAL    H2  -179.991617   60.851468    1.080264   C2    C1    C6
ICOOR_INTERNAL    H1  -179.993980   58.418820    1.080282   C1    C5    C2
ICOOR_INTERNAL    C27  179.982858   59.552969    1.520825   C13   C10   C5
ICOOR_INTERNAL    F1  -179.987880   67.606252    1.384916   C27   C13   C10
ICOOR_INTERNAL    F2   120.054389   70.477005    1.383014   C27   C13   F1
ICOOR_INTERNAL    F3   119.904788   70.480419    1.383074   C27   C13   F2
ICOOR_INTERNAL    H5  -179.992947   59.638075    1.080279   C10   C15   C13
ICOOR_INTERNAL    H9  -179.990293   62.633293    1.087843   C19   C20   C15
ICOOR_INTERNAL    H10  179.985237   61.090752    1.080489   C20   C18   C19
ICOOR_INTERNAL    H8  -179.976657   62.342554    1.087064   C18   C9    C20
ICOOR_INTERNAL    N1   179.169058   66.367720    1.317939   C9    S1    C18
ICOOR_INTERNAL    H4   178.050628   50.571171    1.080506   C4    C8    S1
ICOOR_INTERNAL    N8   179.999349   66.189768    1.468382   C21   C8    C4
ICOOR_INTERNAL    S2  -179.028528   58.301695    1.649480   N8    C21   C8
ICOOR_INTERNAL    O2   -83.828030   72.115214    1.437728   S2    N8    C21
ICOOR_INTERNAL    O3  -110.979907   68.095120    1.445125   S2    N8    O2
ICOOR_INTERNAL    O8  -129.796485   67.884564    1.508898   S2    N8    O3
ICOOR_INTERNAL    C26   47.972804   59.847207    1.411561   O8    S2    N8
ICOOR_INTERNAL    C24  157.660576   70.738957    1.511080   C26   O8    S2
ICOOR_INTERNAL    C22 -162.092351   70.936507    1.469262   C24   C26   O8
ICOOR_INTERNAL    C23 -144.249833   77.197932    1.465575   C22   C24   C26
ICOOR_INTERNAL    C25   31.366047   77.527376    1.475916   C23   C22   C24
ICOOR_INTERNAL    N5    91.982349   65.429978    1.445332   C25   C23   C22
ICOOR_INTERNAL    C12  -93.611780   52.728629    1.352156   N5    C25   C23
ICOOR_INTERNAL    N4  -179.129471   69.965230    1.324715   C12   N5    C25
ICOOR_INTERNAL    C14    0.433861   70.203038    1.328601   N4    C12   N5
ICOOR_INTERNAL    C16   -0.110422   73.543260    1.420679   C14   N4    C12
ICOOR_INTERNAL    N3   179.726204   60.629497    1.342216   C16   C14   N4
ICOOR_INTERNAL    C11    0.608176   59.624923    1.326020   N3    C16   C14
ICOOR_INTERNAL    N2    -0.848435   57.610071    1.331554   C11   N3    C16
ICOOR_INTERNAL    C17    0.552053   58.120141    1.342096   N2    C11   N3
ICOOR_INTERNAL    N7  -179.890730   60.506917    1.449335   C17   N2    C11
ICOOR_INTERNAL    H19  179.859865   60.001164    0.984524   N7    C17   N2
ICOOR_INTERNAL    H20  179.991250   60.000412    0.984519   N7    C17   H19
ICOOR_INTERNAL    H6  -179.999269   61.192473    1.031974   C11   N3    N2
ICOOR_INTERNAL    H7   179.997573   55.017697    1.031995   C12   N5    N4
ICOOR_INTERNAL    O5  -122.223815   71.733653    1.413900   C25   C23   N5
ICOOR_INTERNAL    H14 -121.130496   73.235732    1.069989   C25   C23   O5
ICOOR_INTERNAL    O7  -116.511459   66.309694    1.374530   C23   C22   C25
ICOOR_INTERNAL    H23 -179.997776   70.530952    0.969950   O7    C23   C22
ICOOR_INTERNAL    H12 -118.778660   69.018066    1.070015   C23   C22   O7
ICOOR_INTERNAL    O6  -122.097254   69.957154    1.380198   C22   C24   C23
ICOOR_INTERNAL    H22  -64.966530   70.526952    0.970044   O6    C22   C24
ICOOR_INTERNAL    H11 -117.021961   64.834299    1.070064   C22   C24   O6
ICOOR_INTERNAL    H13  120.000334   70.466931    1.069977   C24   C26   C22
ICOOR_INTERNAL    H15 -119.967044   70.476305    1.069971   C26   O8    C24
ICOOR_INTERNAL    H16 -119.934519   70.408073    1.070003   C26   O8    H15
```

```
ICOOR_INTERNAL    H21 -179.999258   60.849549    0.984516   N8    C21   S2
ICOOR_INTERNAL    O1  -177.630746   57.240813    1.227206   C21   C8    N8
PDB_ROTAMERS LCCa_conformers.pdb
```

CouLuc-2-NEt_2 params file was as follows.

```
NAME LCG
IO_STRING LCG Z
TYPE LIGAND
AA UNK
ATOM  C20 CNH2  X   0.81
ATOM  C16 CH1   X   0.21
ATOM  C8  aroC  X   0.08
ATOM  S1  S     X   -0.28
ATOM  C15 aroC  X   0.58
ATOM  C19 aroC  X   -0.14
ATOM  C18 aroC  X   -0.15
ATOM  C11 aroC  X   0.01
ATOM  C7  aroC  X   -0.18
ATOM  C4  aroC  X   0.03
ATOM  C1  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C5  aroC  X   0.10
ATOM  C3  aroC  X   -0.15
ATOM  C6  aroC  X   0.08
ATOM  O5  Oaro  X   -0.23
ATOM  C14 COO   X   0.71
ATOM  O1  OOC   X   -0.57
ATOM  H3  Haro  X   0.15
ATOM  N8  Nhis  X   -0.84
ATOM  C28 CH2   X   0.37
ATOM  C25 CH3   X   0.00
ATOM  H14 Hapo  X   0.00
ATOM  H15 Hapo  X   0.00
ATOM  H16 Hapo  X   0.00
ATOM  H22 Hapo  X   0.00
ATOM  H23 Hapo  X   0.00
ATOM  C29 CH2   X   0.37
ATOM  C26 CH3   X   0.00
ATOM  H17 Hapo  X   0.00
ATOM  H18 Hapo  X   0.00
ATOM  H19 Hapo  X   0.00
ATOM  H24 Hapo  X   0.00
ATOM  H25 Hapo  X   0.00
ATOM  H2  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  H4  Haro  X   0.15
ATOM  H8  Haro  X   0.15
ATOM  H9  Haro  X   0.15
ATOM  N4  Nhis  X   -0.66
ATOM  H5  Haro  X   0.06
ATOM  N7  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O3  OOC   X   -0.65
ATOM  O4  OOC   X   -0.65
```

```
ATOM  O9  OH    X   -0.46
ATOM  C27 CH2   X   0.28
ATOM  C23 CH1   X   0.28
ATOM  C21 CH1   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C24 CH1   X   0.63
ATOM  N5  Npro  X   -0.08
ATOM  C10 aroC  X   0.55
ATOM  N3  Nhis  X   -0.66
ATOM  C12 aroC  X   0.46
ATOM  C13 aroC  X   0.78
ATOM  N2  Nhis  X   -0.66
ATOM  C9  aroC  X   0.60
ATOM  N1  Nhis  X   -0.65
ATOM  C17 CH1   X   0.51
ATOM  N6  NH2O  X   -0.12
ATOM  H26 Hpol  X   0.40
ATOM  H27 Hpol  X   0.40
ATOM  H6  Haro  X   0.06
ATOM  H7  Haro  X   0.06
ATOM  O6  OH    X   -0.56
ATOM  H13 Hapo  X   0.00
ATOM  O8  OH    X   -0.68
ATOM  H30 Hpol  X   0.40
ATOM  H11 Hapo  X   0.00
ATOM  O7  OH    X   -0.68
ATOM  H29 Hpol  X   0.40
ATOM  H10 Hapo  X   0.00
ATOM  H12 Hapo  X   0.00
ATOM  H20 Hapo  X   0.00
ATOM  H21 Hapo  X   0.00
ATOM  H28 Hpol  X   0.42
ATOM  O2  ONH2  X   -0.57
BOND_TYPE  C1   C2  4
BOND_TYPE  C1   C4  4
BOND_TYPE  C2   C5  4
BOND_TYPE  C3   C5  4
BOND_TYPE  C3   C6  4
BOND_TYPE  C4   C6  4
BOND_TYPE  C4   C7  1
BOND_TYPE  C5   N8  1
BOND_TYPE  C6   O5  1
BOND_TYPE  C7   C11 2
BOND_TYPE  C8   C16 1
BOND_TYPE  C8   S1  2
BOND_TYPE  C9   N1  2
BOND_TYPE  C9   N2  1
BOND_TYPE  C10  N3  1
BOND_TYPE  C10  N5  2
BOND_TYPE  C11  C14 1
BOND_TYPE  C11  C18 1
BOND_TYPE  C12  C13 1
BOND_TYPE  C12  C17 1
BOND_TYPE  C12  N3  2
BOND_TYPE  C13  N2  2
```

```
BOND_TYPE  C13  N5  1
BOND_TYPE  C14  O1  2
BOND_TYPE  C14  O5  1
BOND_TYPE  C15  C19 1
BOND_TYPE  C15  N4  2
BOND_TYPE  C15  S1  1
BOND_TYPE  C16  C20 1
BOND_TYPE  C16  N4  1
BOND_TYPE  C17  N1  1
BOND_TYPE  C17  N6  1
BOND_TYPE  C18  C19 2
BOND_TYPE  C20  N7  4
BOND_TYPE  C20  O2  2
BOND_TYPE  C21  C22 1
BOND_TYPE  C21  C23 1
BOND_TYPE  C21  O7  1
BOND_TYPE  C22  C24 1
BOND_TYPE  C22  O8  1
BOND_TYPE  C23  C27 1
BOND_TYPE  C23  O6  1
BOND_TYPE  C24  N5  1
BOND_TYPE  C24  O6  1
BOND_TYPE  C25  C28 1
BOND_TYPE  C26  C29 1
BOND_TYPE  C27  O9  1
BOND_TYPE  C28  N8  1
BOND_TYPE  C29  N8  1
BOND_TYPE  N7   S2  1
BOND_TYPE  O3   S2  2
BOND_TYPE  O4   S2  2
BOND_TYPE  O9   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C7   H4  1
BOND_TYPE  C8   H5  1
BOND_TYPE  C9   H6  1
BOND_TYPE  C10  H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C19  H9  1
BOND_TYPE  C21  H10 1
BOND_TYPE  C22  H11 1
BOND_TYPE  C23  H12 1
BOND_TYPE  C24  H13 1
BOND_TYPE  C25  H14 1
BOND_TYPE  C25  H15 1
BOND_TYPE  C25  H16 1
BOND_TYPE  C26  H17 1
BOND_TYPE  C26  H18 1
BOND_TYPE  C26  H19 1
BOND_TYPE  C27  H20 1
BOND_TYPE  C27  H21 1
BOND_TYPE  C28  H22 1
BOND_TYPE  C28  H23 1
BOND_TYPE  C29  H24 1
```

```
BOND_TYPE  C29  H25 1
BOND_TYPE  N6   H26 1
BOND_TYPE  N6   H27 1
BOND_TYPE  N7   H28 1
BOND_TYPE  O7   H29 1
BOND_TYPE  O8   H30 1
CHI 1  C23  C21  O7   H29
#PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 2  C21  C22  O8   H30
#PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 3  C2   C5   N8   C28
#PROTON_CHI 3 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 4  C19  C18  C11  C7
#PROTON_CHI 4 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 5  S1   C15  C19  C18
#PROTON_CHI 5 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 6  N7   C20  C16  C8
CHI 7  O9   C27  C23  C21
CHI 8  C22  C24  N5   C10
CHI 9  S2   O9   C27  C23
CHI 10  C5   N8   C28  C25
#PROTON_CHI 10 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 11  C5   N8   C29  C26
#PROTON_CHI 11 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 12  C20  N7   S2   O3
CHI 13  N7   S2   O9   C27
NBR_ATOM  C20
NBR_RADIUS 18.224748
ICOOR_INTERNAL    C20    0.000000    0.000000    0.000000   C20   C16
C8
ICOOR_INTERNAL    C16    0.000000  180.000000    1.496021   C20   C16
C8
ICOOR_INTERNAL    C8     0.000000   59.872347    1.372318   C16   C20
C8
ICOOR_INTERNAL    S1  -179.270493   77.666388    1.751132   C8    C16
C20
ICOOR_INTERNAL    C15    0.492363   84.355877    1.747855   S1    C8
C16
ICOOR_INTERNAL    C19 -178.965358   51.788039    1.472562   C15   S1
C8
ICOOR_INTERNAL    C18   -1.305325   59.030868    1.347370   C19   C15
S1
ICOOR_INTERNAL    C11 -179.274860   53.166278    1.497084   C18   C19
C15
ICOOR_INTERNAL    C7   179.451266   63.097829    1.404762   C11   C18
C19
ICOOR_INTERNAL    C4  -179.793688   59.532570    1.395584   C7    C11
C18
ICOOR_INTERNAL    C1   179.796054   58.392558    1.400612   C4    C7
C11
ICOOR_INTERNAL    C2  -179.921545   60.040291    1.397529   C1    C4
C7
ICOOR_INTERNAL    C5    -0.116957   59.289576    1.406615   C2    C1
C4
```

```
ICOOR_INTERNAL    C3     0.075829   61.084161    1.406503   C5    C2
C1
ICOOR_INTERNAL    C6     0.052323   59.387740    1.398536   C3    C5
C2
ICOOR_INTERNAL    O5   179.736043   60.042539    1.327524   C6    C3
C5
ICOOR_INTERNAL    C14 -179.964796   56.803196    1.330734   O5    C6
C3
ICOOR_INTERNAL    O1  -179.927603   61.884600    1.215817   C14   O5
C6
ICOOR_INTERNAL    H3   178.842523   59.779401    1.082402   C3    C5
C6
ICOOR_INTERNAL    N8  -179.993686   59.454771    1.434979   C5    C2
C3
ICOOR_INTERNAL    C28   -0.002457   57.690884    1.466778   N8    C5
C2
ICOOR_INTERNAL    C25  -91.854520   64.095211    1.534414   C28   N8
C5
ICOOR_INTERNAL    H14   60.685126   69.691289    1.110806   C25   C28
N8
ICOOR_INTERNAL    H15 -118.802735   71.019176    1.110603   C25   C28
H14
ICOOR_INTERNAL    H16 -118.056018   67.538937    1.105051   C25   C28
H15
ICOOR_INTERNAL    H22 -118.146161   72.185330    1.112371   C28   N8
C25
ICOOR_INTERNAL    H23 -117.451347   70.323011    1.111477   C28   N8
H22
ICOOR_INTERNAL    C29 -179.995647   57.678678    1.466716   N8    C5
C28
ICOOR_INTERNAL    C26  -91.850411   64.105071    1.534379   C29   N8
C5
ICOOR_INTERNAL    H17   60.685925   67.549597    1.105211   C26   C29
N8
ICOOR_INTERNAL    H18 -123.147627   69.685529    1.110768   C26   C29
H17
ICOOR_INTERNAL    H19 -118.811804   71.020849    1.110601   C26   C29
H18
ICOOR_INTERNAL    H24 -118.150108   72.182516    1.112383   C29   N8
C26
ICOOR_INTERNAL    H25 -117.454404   70.317350    1.111475   C29   N8
H24
ICOOR_INTERNAL    H2  -178.850362   61.102830    1.081904   C2    C1
C5
ICOOR_INTERNAL    H1  -179.822022   59.308872    1.083577   C1    C4
C2
ICOOR_INTERNAL    H4   179.822139   59.822077    1.084739   C7    C11
C4
ICOOR_INTERNAL    H8   179.434073   62.837482    1.087641   C18   C19
C11
ICOOR_INTERNAL    H9  -179.278443   63.724142    1.078562   C19   C15
C18
ICOOR_INTERNAL    N4   178.910262   74.692820    1.329868   C15   S1
C19
```

```
ICOOR_INTERNAL    H5   179.269279   46.636101    1.077222   C8    C16
S1
ICOOR_INTERNAL    N7    -9.923225   63.813358    1.468378   C20   C16
C8
ICOOR_INTERNAL    S2   178.603099   58.300039    1.649518   N7    C20
C16
ICOOR_INTERNAL    O3   -83.829480   72.117504    1.437768   S2    N7
C20
ICOOR_INTERNAL    O4  -110.978218   68.095528    1.445070   S2    N7
O3
ICOOR_INTERNAL    O9  -129.803411   67.884839    1.508885   S2    N7
O4
ICOOR_INTERNAL    C27   47.977324   59.848471    1.411546   O9    S2
N7
ICOOR_INTERNAL    C23  157.655249   70.737770    1.511083   C27   O9
S2
ICOOR_INTERNAL    C21 -162.093482   70.933727    1.469270   C23   C27
O9
ICOOR_INTERNAL    C22 -144.247783   77.196739    1.465552   C21   C23
C27
ICOOR_INTERNAL    C24   31.367605   77.529154    1.475957   C22   C21
C23
ICOOR_INTERNAL    N5    91.983558   65.428300    1.445328   C24   C22
C21
ICOOR_INTERNAL    C10  -93.613244   52.728471    1.352163   N5    C24
C22
ICOOR_INTERNAL    N3  -179.130300   69.967367    1.324677   C10   N5
C24
ICOOR_INTERNAL    C12    0.433704   70.197497    1.328561   N3    C10
N5
ICOOR_INTERNAL    C13   -0.110640   73.547282    1.420761   C12   N3
C10
ICOOR_INTERNAL    N2   179.726171   60.632752    1.342181   C13   C12
N3
ICOOR_INTERNAL    C9     0.608583   59.621629    1.326036   N2    C13
C12
ICOOR_INTERNAL    N1    -0.848335   57.613744    1.331577   C9    N2
C13
ICOOR_INTERNAL    C17    0.553656   58.116069    1.342041   N1    C9
N2
ICOOR_INTERNAL    N6  -179.889752   60.503702    1.449271   C17   N1
C9
ICOOR_INTERNAL    H26 -152.990791   73.020370    1.019094   N6    C17
N1
ICOOR_INTERNAL    H27  125.651839   73.003358    1.011416   N6    C17
H26
ICOOR_INTERNAL    H6  -179.346779   61.198218    1.081267   C9    N2
N1
ICOOR_INTERNAL    H7   176.292279   47.119880    1.018607   C10   N5
N3
ICOOR_INTERNAL    O6  -122.227965   71.734996    1.413847   C24   C22
N5
ICOOR_INTERNAL    H13 -118.377782   70.274077    1.114966   C24   C22
O6
```

```
ICOOR_INTERNAL    O8  -116.514168   66.311108    1.374557   C22   C21
C24
ICOOR_INTERNAL    H30 -179.998770   69.489626    0.994258   O8    C22
C21
ICOOR_INTERNAL    H11 -120.121966   66.508026    1.112708   C22   C21
O8
ICOOR_INTERNAL    O7  -122.099284   69.960468    1.380236   C21   C23
C22
ICOOR_INTERNAL    H29  -64.969184   73.081754    0.992629   O7    C21
C23
ICOOR_INTERNAL    H10 -126.339895   68.634703    1.120794   C21   C23
O7
ICOOR_INTERNAL    H12  120.013706   68.521390    1.115529   C23   C27
C21
ICOOR_INTERNAL    H20 -117.558614   72.406306    1.115829   C27   O9
C23
ICOOR_INTERNAL    H21 -120.541374   68.798568    1.089626   C27   O9
H20
ICOOR_INTERNAL    H28  179.382356   56.641763    1.023812   N7    C20
S2
ICOOR_INTERNAL    O2  -179.988605   59.568862    1.227216   C20   C16
N7
```

The NapLuc-2-NMe_2 params file was as follows.

```
NAME NLA
IO_STRING NLA Z
TYPE LIGAND
AA UNK
ATOM  C21 CNH2  X   0.89
ATOM  C12 aroC  X   0.14
ATOM  C7  aroC  X   -0.11
ATOM  S1  S     X   -0.08
ATOM  C13 aroC  X   0.33
ATOM  C20 aroC  X   -0.10
ATOM  C19 aroC  X   -0.18
ATOM  C10 aroC  X   0.03
ATOM  C3  aroC  X   -0.15
ATOM  C1  aroC  X   -0.15
ATOM  C9  aroC  X   0.00
ATOM  C6  aroC  X   -0.15
ATOM  C11 aroC  X   0.10
ATOM  C4  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C8  aroC  X   0.00
ATOM  C5  aroC  X   -0.15
ATOM  H5  Haro  X   0.15
ATOM  H2  Haro  X   0.15
ATOM  H4  Haro  X   0.15
ATOM  N8  Nhis  X   -0.84
ATOM  C26 CH3   X   0.37
ATOM  H16 Hapo  X   0.00
ATOM  H17 Hapo  X   0.00
ATOM  H18 Hapo  X   0.00
ATOM  C27 CH3   X   0.37
```

```
ATOM  H19 Hapo  X   0.00
ATOM  H20 Hapo  X   0.00
ATOM  H21 Hapo  X   0.00
ATOM  H6  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  H3  Haro  X   0.15
ATOM  H10 Haro  X   0.15
ATOM  H11 Haro  X   0.15
ATOM  N1  Nhis  X   -0.57
ATOM  H7  Haro  X   0.15
ATOM  N7  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O2  OOC   X   -0.65
ATOM  O3  OOC   X   -0.65
ATOM  O7  OH    X   -0.46
ATOM  C28 CH2   X   0.28
ATOM  C24 CH1   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C23 CH1   X   0.28
ATOM  C25 CH1   X   0.63
ATOM  N5  Npro  X   -0.08
ATOM  C15 aroC  X   0.55
ATOM  N4  Nhis  X   -0.66
ATOM  C16 aroC  X   0.46
ATOM  C17 aroC  X   0.78
ATOM  N3  Nhis  X   -0.66
ATOM  C14 aroC  X   0.60
ATOM  N2  Nhis  X   -0.65
ATOM  C18 CH1   X   0.51
ATOM  N6  NH2O  X   -0.12
ATOM  H24 Hpol  X   0.40
ATOM  H25 Hpol  X   0.40
ATOM  H8  Haro  X   0.06
ATOM  H9  Haro  X   0.06
ATOM  O4  OH    X   -0.56
ATOM  H15 Hapo  X   0.00
ATOM  O6  OH    X   -0.68
ATOM  H28 Hpol  X   0.40
ATOM  H13 Hapo  X   0.00
ATOM  O5  OH    X   -0.68
ATOM  H27 Hpol  X   0.40
ATOM  H12 Hapo  X   0.00
ATOM  H14 Hapo  X   0.00
ATOM  H22 Hapo  X   0.00
ATOM  H23 Hapo  X   0.00
ATOM  H26 Hpol  X   0.42
ATOM  O1  ONH2  X   -0.57
BOND_TYPE  C1   C3  4
BOND_TYPE  C1   C9  4
BOND_TYPE  C2   C4  4
BOND_TYPE  C2   C8  4
BOND_TYPE  C3   C10 4
BOND_TYPE  C4   C11 4
BOND_TYPE  C5   C8  4
BOND_TYPE  C5   C10 4
```

```
BOND_TYPE  C6   C9  4
BOND_TYPE  C6   C11 4
BOND_TYPE  C7   C12 2
BOND_TYPE  C7   S1  1
BOND_TYPE  C8   C9  4
BOND_TYPE  C10  C19 1
BOND_TYPE  C11  N8  1
BOND_TYPE  C12  C21 1
BOND_TYPE  C12  N1  1
BOND_TYPE  C13  C20 1
BOND_TYPE  C13  N1  2
BOND_TYPE  C13  S1  1
BOND_TYPE  C14  N2  2
BOND_TYPE  C14  N3  1
BOND_TYPE  C15  N4  1
BOND_TYPE  C15  N5  2
BOND_TYPE  C16  C17 1
BOND_TYPE  C16  C18 1
BOND_TYPE  C16  N4  2
BOND_TYPE  C17  N3  2
BOND_TYPE  C17  N5  1
BOND_TYPE  C18  N2  1
BOND_TYPE  C18  N6  1
BOND_TYPE  C19  C20 2
BOND_TYPE  C21  N7  4
BOND_TYPE  C21  O1  2
BOND_TYPE  C22  C23 1
BOND_TYPE  C22  C24 1
BOND_TYPE  C22  O5  1
BOND_TYPE  C23  C25 1
BOND_TYPE  C23  O6  1
BOND_TYPE  C24  C28 1
BOND_TYPE  C24  O4  1
BOND_TYPE  C25  N5  1
BOND_TYPE  C25  O4  1
BOND_TYPE  C26  N8  1
BOND_TYPE  C27  N8  1
BOND_TYPE  C28  O7  1
BOND_TYPE  N7   S2  1
BOND_TYPE  O2   S2  2
BOND_TYPE  O3   S2  2
BOND_TYPE  O7   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C4   H4  1
BOND_TYPE  C5   H5  1
BOND_TYPE  C6   H6  1
BOND_TYPE  C7   H7  1
BOND_TYPE  C14  H8  1
BOND_TYPE  C15  H9  1
BOND_TYPE  C19  H10 1
BOND_TYPE  C20  H11 1
BOND_TYPE  C22  H12 1
BOND_TYPE  C23  H13 1
```

```
BOND_TYPE  C24  H14 1
BOND_TYPE  C25  H15 1
BOND_TYPE  C26  H16 1
BOND_TYPE  C26  H17 1
BOND_TYPE  C26  H18 1
BOND_TYPE  C27  H19 1
BOND_TYPE  C27  H20 1
BOND_TYPE  C27  H21 1
BOND_TYPE  C28  H22 1
BOND_TYPE  C28  H23 1
BOND_TYPE  N6   H24 1
BOND_TYPE  N6   H25 1
BOND_TYPE  N7   H26 1
BOND_TYPE  O5   H27 1
BOND_TYPE  O6   H28 1
CHI 1  C24  C22  O5   H27
#PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 2  C22  C23  O6   H28
#PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 3  C20  C19  C10  C3
#PROTON_CHI 3 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 4  C6   C11  N8   C26
#PROTON_CHI 4 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 5  N7   C21  C12  C7
CHI 6  S1   C13  C20  C19
#PROTON_CHI 6 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 7  O7   C28  C24  C22
CHI 8  C23  C25  N5   C15
CHI 9  S2   O7   C28  C24
CHI 10  C21  N7   S2   O2
CHI 11  N7   S2   O7   C28
NBR_ATOM  C21
NBR_RADIUS 16.998837
ICOOR_INTERNAL    C21    0.000000    0.000000    0.000000   C21   C12
C7
ICOOR_INTERNAL    C12    0.000000  180.000000    1.496021   C21   C12
C7
ICOOR_INTERNAL    C7     0.000000   59.872347    1.372318   C12   C21
C7
ICOOR_INTERNAL    S1  -179.270493   77.666388    1.751132   C7    C12
C21
ICOOR_INTERNAL    C13    0.492363   84.355877    1.747855   S1    C7
C12
ICOOR_INTERNAL    C20 -173.771619   61.683298    1.726380   C13   S1
C7
ICOOR_INTERNAL    C19  171.268212   57.214464    1.342253   C20   C13
S1
ICOOR_INTERNAL    C10  179.597277   55.226679    1.489884   C19   C20
C13
ICOOR_INTERNAL    C3  -160.002537   61.928409    1.404869   C10   C19
C20
ICOOR_INTERNAL    C1  -179.968581   59.087473    1.396545   C3    C10
C19
ICOOR_INTERNAL    C9    -0.006186   59.677313    1.401283   C1    C3
C10
```

```
ICOOR_INTERNAL    C6   179.987822   59.415405    1.403816   C9    C1
C3
ICOOR_INTERNAL    C11 -179.987152   58.846339    1.414986   C6    C9
C1
ICOOR_INTERNAL    C4    -0.006973   62.221683    1.413997   C11   C6
C9
ICOOR_INTERNAL    C2     0.010676   58.765899    1.396493   C4    C11
C6
ICOOR_INTERNAL    C8    -0.008063   59.529665    1.399099   C2    C4
C11
ICOOR_INTERNAL    C5   179.985804   59.390822    1.404579   C8    C2
C4
ICOOR_INTERNAL    H5     0.004226   61.554529    1.080184   C5    C8
C2
ICOOR_INTERNAL    H2  -179.996492   61.005780    1.083803   C2    C4
C8
ICOOR_INTERNAL    H4   179.988820   58.440586    1.080434   C4    C11
C2
ICOOR_INTERNAL    N8   179.998366   58.840031    1.445016   C11   C6
C4
ICOOR_INTERNAL    C26 -179.996136   59.372524    1.457096   N8    C11
C6
ICOOR_INTERNAL    H16  179.999819   67.581418    1.110235   C26   N8
C11
ICOOR_INTERNAL    H17 -119.548441   70.125007    1.110555   C26   N8
H16
ICOOR_INTERNAL    H18 -120.914948   70.131366    1.110560   C26   N8
H17
ICOOR_INTERNAL    C27  179.987400   59.307240    1.456978   N8    C11
C26
ICOOR_INTERNAL    H19  179.999960   70.119296    1.110451   C27   N8
C11
ICOOR_INTERNAL    H20 -119.532614   67.589575    1.110170   C27   N8
H19
ICOOR_INTERNAL    H21 -119.522494   70.131372    1.110635   C27   N8
H20
ICOOR_INTERNAL    H6   179.997205   61.890803    1.081476   C6    C9
C11
ICOOR_INTERNAL    H1  -179.993733   60.976718    1.083611   C1    C3
C9
ICOOR_INTERNAL    H3   179.982726   59.580568    1.083325   C3    C10
C1
ICOOR_INTERNAL    H10  179.898474   62.179240    1.084956   C19   C20
C10
ICOOR_INTERNAL    H11  179.205681   66.627727    1.084284   C20   C13
C19
ICOOR_INTERNAL    N1   173.716522   74.692820    1.329868   C13   S1
C20
ICOOR_INTERNAL    H7   179.269279   46.636101    1.077222   C7    C12
S1
ICOOR_INTERNAL    N7    -9.923225   63.813358    1.468378   C21   C12
C7
ICOOR_INTERNAL    S2   178.603099   58.300039    1.649518   N7    C21
C12
```

```
ICOOR_INTERNAL    O2   -83.829480   72.117504    1.437768   S2    N7
C21
ICOOR_INTERNAL    O3  -110.978218   68.095528    1.445070   S2    N7
O2
ICOOR_INTERNAL    O7  -129.803411   67.884839    1.508885   S2    N7
O3
ICOOR_INTERNAL    C28   47.977324   59.848471    1.411546   O7    S2
N7
ICOOR_INTERNAL    C24  157.655249   70.737770    1.511083   C28   O7
S2
ICOOR_INTERNAL    C22 -162.093482   70.933727    1.469270   C24   C28
O7
ICOOR_INTERNAL    C23 -144.247783   77.196739    1.465552   C22   C24
C28
ICOOR_INTERNAL    C25   31.367605   77.529154    1.475957   C23   C22
C24
ICOOR_INTERNAL    N5    91.983558   65.428300    1.445328   C25   C23
C22
ICOOR_INTERNAL    C15  -93.613244   52.728471    1.352163   N5    C25
C23
ICOOR_INTERNAL    N4  -179.130300   69.967367    1.324677   C15   N5
C25
ICOOR_INTERNAL    C16    0.433704   70.197497    1.328561   N4    C15
N5
ICOOR_INTERNAL    C17   -0.110640   73.547282    1.420761   C16   N4
C15
ICOOR_INTERNAL    N3   179.726171   60.632752    1.342181   C17   C16
N4
ICOOR_INTERNAL    C14    0.608583   59.621629    1.326036   N3    C17
C16
ICOOR_INTERNAL    N2    -0.848335   57.613744    1.331577   C14   N3
C17
ICOOR_INTERNAL    C18    0.553656   58.116069    1.342041   N2    C14
N3
ICOOR_INTERNAL    N6  -179.889752   60.503702    1.449271   C18   N2
C14
ICOOR_INTERNAL    H24 -152.990791   73.020370    1.019094   N6    C18
N2
ICOOR_INTERNAL    H25  125.651839   73.003358    1.011416   N6    C18
H24
ICOOR_INTERNAL    H8  -179.346779   61.198218    1.081267   C14   N3
N2
ICOOR_INTERNAL    H9   176.292279   47.119880    1.018607   C15   N5
N4
ICOOR_INTERNAL    O4  -122.227965   71.734996    1.413847   C25   C23
N5
ICOOR_INTERNAL    H15 -118.377782   70.274077    1.114966   C25   C23
O4
ICOOR_INTERNAL    O6  -116.514168   66.311108    1.374557   C23   C22
C25
ICOOR_INTERNAL    H28 -179.998770   69.489626    0.994258   O6    C23
C22
ICOOR_INTERNAL    H13 -120.121966   66.508026    1.112708   C23   C22
O6
```

```
ICOOR_INTERNAL    O5  -122.099284   69.960468    1.380236   C22   C24
C23
ICOOR_INTERNAL    H27  -64.969184   73.081754    0.992629   O5    C22
C24
ICOOR_INTERNAL    H12 -126.339895   68.634703    1.120794   C22   C24
O5
ICOOR_INTERNAL    H14  120.013706   68.521390    1.115529   C24   C28
C22
ICOOR_INTERNAL    H22 -117.558614   72.406306    1.115829   C28   O7
C24
ICOOR_INTERNAL    H23 -120.541374   68.798568    1.089626   C28   O7
H22
ICOOR_INTERNAL    H26  179.382356   56.641763    1.023812   N7    C21
S2
ICOOR_INTERNAL    O1  -179.988605   59.568862    1.227216   C21   C12
N7
```

FPLuc 1b

```
NAME RSB
IO_STRING RSB Z
TYPE LIGAND
AA UNK
ATOM  C19 CNH2  X   0.81
ATOM  C16 CH1   X   0.21
ATOM  C7  aroC  X   0.08
ATOM  S1  S     X   -0.28
ATOM  C14 aroC  X   0.59
ATOM  C12 aroC  X   0.51
ATOM  N4  Nhis  X   -0.62
ATOM  C11 aroC  X   0.19
ATOM  C15 CNH2  X   0.62
ATOM  N6  Npro  X   -0.42
ATOM  C24 CH3   X   0.30
ATOM  H13 Hapo  X   0.00
ATOM  H14 Hapo  X   0.00
ATOM  H15 Hapo  X   0.00
ATOM  O1  ONH2  X   -0.57
ATOM  C18 aroC  X   -0.18
ATOM  C5  aroC  X   0.03
ATOM  C1  aroC  X   -0.15
ATOM  C3  aroC  X   -0.15
ATOM  C6  aroC  X   0.10
ATOM  C4  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  H2  Haro  X   0.15
ATOM  H4  Haro  X   0.15
ATOM  N10 Nhis  X   -0.84
ATOM  C25 CH3   X   0.37
ATOM  H16 Hapo  X   0.00
ATOM  H17 Hapo  X   0.00
ATOM  H18 Hapo  X   0.00
```

```
ATOM  C26 CH3   X   0.37
ATOM  H19 Hapo  X   0.00
ATOM  H20 Hapo  X   0.00
ATOM  H21 Hapo  X   0.00
ATOM  H3  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  H8  Haro  X   0.15
ATOM  N5  Nhis  X   -0.66
ATOM  H5  Haro  X   0.06
ATOM  N9  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O3  OOC   X   -0.65
ATOM  O4  OOC   X   -0.65
ATOM  O8  OH    X   -0.46
ATOM  C27 CH2   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C20 CH1   X   0.28
ATOM  C21 CH1   X   0.28
ATOM  C23 CH1   X   0.63
ATOM  N7  Npro  X   -0.08
ATOM  C9  aroC  X   0.55
ATOM  N3  Nhis  X   -0.66
ATOM  C10 aroC  X   0.46
ATOM  C13 aroC  X   0.78
ATOM  N2  Nhis  X   -0.66
ATOM  C8  aroC  X   0.60
ATOM  N1  Nhis  X   -0.65
ATOM  C17 CH1   X   0.51
ATOM  N8  NH2O  X   -0.12
ATOM  H24 Hpol  X   0.40
ATOM  H25 Hpol  X   0.40
ATOM  H6  Haro  X   0.06
ATOM  H7  Haro  X   0.06
ATOM  O5  OH    X   -0.56
ATOM  H12 Hapo  X   0.00
ATOM  O7  OH    X   -0.68
ATOM  H28 Hpol  X   0.40
ATOM  H10 Hapo  X   0.00
ATOM  O6  OH    X   -0.68
ATOM  H27 Hpol  X   0.40
ATOM  H9  Hapo  X   0.00
ATOM  H11 Hapo  X   0.00
ATOM  H22 Hapo  X   0.00
ATOM  H23 Hapo  X   0.00
ATOM  H26 Hpol  X   0.42
ATOM  O2  ONH2  X   -0.57
BOND_TYPE  C1   C3  4
BOND_TYPE  C1   C5  4
BOND_TYPE  C2   C4  4
BOND_TYPE  C2   C5  4
BOND_TYPE  C3   C6  4
BOND_TYPE  C4   C6  4
BOND_TYPE  C5   C18 1
BOND_TYPE  C6   N10 1
BOND_TYPE  C7   C16 1
```

```
BOND_TYPE  C7   S1  2
BOND_TYPE  C8   N1  2
BOND_TYPE  C8   N2  1
BOND_TYPE  C9   N3  1
BOND_TYPE  C9   N7  2
BOND_TYPE  C10  C13 1
BOND_TYPE  C10  C17 1
BOND_TYPE  C10  N3  2
BOND_TYPE  C11  C15 1
BOND_TYPE  C11  C18 2
BOND_TYPE  C11  N4  1
BOND_TYPE  C12  C14 1
BOND_TYPE  C12  N4  2
BOND_TYPE  C12  N6  1
BOND_TYPE  C13  N2  2
BOND_TYPE  C13  N7  1
BOND_TYPE  C14  N5  2
BOND_TYPE  C14  S1  1
BOND_TYPE  C15  N6  4
BOND_TYPE  C15  O1  2
BOND_TYPE  C16  C19 1
BOND_TYPE  C16  N5  1
BOND_TYPE  C17  N1  1
BOND_TYPE  C17  N8  1
BOND_TYPE  C19  N9  4
BOND_TYPE  C19  O2  2
BOND_TYPE  C20  C21 1
BOND_TYPE  C20  C22 1
BOND_TYPE  C20  O6  1
BOND_TYPE  C21  C23 1
BOND_TYPE  C21  O7  1
BOND_TYPE  C22  C27 1
BOND_TYPE  C22  O5  1
BOND_TYPE  C23  N7  1
BOND_TYPE  C23  O5  1
BOND_TYPE  C24  N6  1
BOND_TYPE  C25  N10 1
BOND_TYPE  C26  N10 1
BOND_TYPE  C27  O8  1
BOND_TYPE  N9   S2  1
BOND_TYPE  O3   S2  2
BOND_TYPE  O4   S2  2
BOND_TYPE  O8   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C4   H4  1
BOND_TYPE  C7   H5  1
BOND_TYPE  C8   H6  1
BOND_TYPE  C9   H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C20  H9  1
BOND_TYPE  C21  H10 1
BOND_TYPE  C22  H11 1
BOND_TYPE  C23  H12 1
```

```
BOND_TYPE  C24  H13 1
BOND_TYPE  C24  H14 1
BOND_TYPE  C24  H15 1
BOND_TYPE  C25  H16 1
BOND_TYPE  C25  H17 1
BOND_TYPE  C25  H18 1
BOND_TYPE  C26  H19 1
BOND_TYPE  C26  H20 1
BOND_TYPE  C26  H21 1
BOND_TYPE  C27  H22 1
BOND_TYPE  C27  H23 1
BOND_TYPE  N8   H24 1
BOND_TYPE  N8   H25 1
BOND_TYPE  N9   H26 1
BOND_TYPE  O6   H27 1
BOND_TYPE  O7   H28 1
CHI 1  C22  C20  O6   H27
PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 2  C20  C21  O7   H28
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 3  C11  C18  C5   C1
CHI 4  C3   C6   N10  C25
CHI 5  S1   C14  C12  N4
CHI 6  N9   C19  C16  C7
CHI 7  O8   C27  C22  C20
CHI 8  C21  C23  N7   C9
CHI 9  S2   O8   C27  C22
CHI 10  C19  N9   S2   O3
CHI 11  N9   S2   O8   C27
NBR_ATOM  C19
NBR_RADIUS 16.914558
ICOOR_INTERNAL    C19    0.000000    0.000000    0.000000   C19   C16
C7
ICOOR_INTERNAL    C16    0.000000  180.000000    1.496095   C19   C16
C7
ICOOR_INTERNAL    C7     0.000000   59.870629    1.372257   C16   C19
C7
ICOOR_INTERNAL    S1  -179.266651   77.664513    1.751119   C7    C16
C19
ICOOR_INTERNAL    C14    0.491524   84.357409    1.747864   S1    C7
C16
ICOOR_INTERNAL    C12  177.286844   49.940206    1.622431   C14   S1
C7
ICOOR_INTERNAL    N4     3.119648   68.967221    1.289347   C12   C14
S1
ICOOR_INTERNAL    C11  179.585573   69.443259    1.404266   N4    C12
C14
ICOOR_INTERNAL    C15    0.217288   73.173984    1.447123   C11   N4
C12
ICOOR_INTERNAL    N6     0.110403   69.777641    1.351244   C15   C11
N4
ICOOR_INTERNAL    C24  178.762000   62.670855    1.536108   N6    C15
C11
ICOOR_INTERNAL    H13  -88.646762   91.959140    1.070045   C24   N6
C15
```

```
ICOOR_INTERNAL    H14 -116.010915   68.624428    1.110449   C24   N6
H13
ICOOR_INTERNAL    H15 -128.473712   68.605163    1.110298   C24   N6
H14
ICOOR_INTERNAL    O1   179.476428   54.497599    1.223134   C15   C11
N6
ICOOR_INTERNAL    C18  179.948502   53.009608    1.348754   C11   N4
C15
ICOOR_INTERNAL    C5    -0.005324   52.277498    1.495927   C18   C11
N4
ICOOR_INTERNAL    C1   119.973951   55.415443    1.406595   C5    C18
C11
ICOOR_INTERNAL    C3   179.939001   59.106594    1.398201   C1    C5
C18
ICOOR_INTERNAL    C6     0.170504   58.470265    1.414210   C3    C1
C5
ICOOR_INTERNAL    C4    -0.075253   63.008161    1.413693   C6    C3
C1
ICOOR_INTERNAL    C2    -0.087128   58.688472    1.396733   C4    C6
C3
ICOOR_INTERNAL    H2  -179.642280   61.602795    1.083640   C2    C4
C6
ICOOR_INTERNAL    H4  -178.616178   57.596882    1.077836   C4    C6
C2
ICOOR_INTERNAL    N10  179.989872   58.488243    1.449368   C6    C3
C4
ICOOR_INTERNAL    C25   -0.005764   57.461630    1.457935   N10   C6
C3
ICOOR_INTERNAL    H16 -179.999783   65.758567    1.102772   C25   N10
C6
ICOOR_INTERNAL    H17 -122.593261   70.093874    1.110752   C25   N10
H16
ICOOR_INTERNAL    H18 -119.299184   70.638298    1.110932   C25   N10
H17
ICOOR_INTERNAL    C26 -179.999514   57.472830    1.457900   N10   C6
C25
ICOOR_INTERNAL    H19  179.998676   65.765352    1.102846   C26   N10
C6
ICOOR_INTERNAL    H20 -122.588868   70.088248    1.110753   C26   N10
H19
ICOOR_INTERNAL    H21 -119.298877   70.638601    1.110919   C26   N10
H20
ICOOR_INTERNAL    H3   178.680828   63.778744    1.077983   C3    C1
C6
ICOOR_INTERNAL    H1   179.779170   57.524698    1.074940   C1    C5
C3
ICOOR_INTERNAL    H8   179.968044   63.933978    1.087264   C18   C11
C5
ICOOR_INTERNAL    N5  -177.338451   74.691871    1.329846   C14   S1
C12
ICOOR_INTERNAL    H5   179.272402   46.637967    1.077202   C7    C16
S1
ICOOR_INTERNAL    N9    -9.928572   63.816343    1.468382   C19   C16
C7
```

```
ICOOR_INTERNAL    S2   178.603415   58.301695    1.649480   N9    C19
C16
ICOOR_INTERNAL    O3   -83.828030   72.115214    1.437728   S2    N9
C19
ICOOR_INTERNAL    O4  -110.979907   68.095120    1.445125   S2    N9
O3
ICOOR_INTERNAL    O8  -129.796485   67.884564    1.508898   S2    N9
O4
ICOOR_INTERNAL    C27   47.972804   59.847207    1.411561   O8    S2
N9
ICOOR_INTERNAL    C22  157.660576   70.738957    1.511080   C27   O8
S2
ICOOR_INTERNAL    C20 -162.092351   70.936507    1.469262   C22   C27
O8
ICOOR_INTERNAL    C21 -144.249833   77.197932    1.465575   C20   C22
C27
ICOOR_INTERNAL    C23   31.366047   77.527376    1.475916   C21   C20
C22
ICOOR_INTERNAL    N7    91.982349   65.429978    1.445332   C23   C21
C20
ICOOR_INTERNAL    C9   -93.611780   52.728629    1.352156   N7    C23
C21
ICOOR_INTERNAL    N3  -179.129471   69.965230    1.324715   C9    N7
C23
ICOOR_INTERNAL    C10    0.433861   70.203038    1.328601   N3    C9
N7
ICOOR_INTERNAL    C13   -0.110422   73.543260    1.420679   C10   N3
C9
ICOOR_INTERNAL    N2   179.726204   60.629497    1.342216   C13   C10
N3
ICOOR_INTERNAL    C8     0.608176   59.624923    1.326020   N2    C13
C10
ICOOR_INTERNAL    N1    -0.848435   57.610071    1.331554   C8    N2
C13
ICOOR_INTERNAL    C17    0.552053   58.120141    1.342096   N1    C8
N2
ICOOR_INTERNAL    N8  -179.890730   60.506917    1.449335   C17   N1
C8
ICOOR_INTERNAL    H24  179.876742   59.668192    1.030159   N8    C17
N1
ICOOR_INTERNAL    H25  179.955977   59.596310    1.030220   N8    C17
H24
ICOOR_INTERNAL    H6  -179.344486   61.198071    1.081258   C8    N2
N1
ICOOR_INTERNAL    H7   177.596100   51.693780    1.047707   C9    N7
N3
ICOOR_INTERNAL    O5  -122.223815   71.733653    1.413900   C23   C21
N7
ICOOR_INTERNAL    H12 -118.383618   70.268459    1.114938   C23   C21
O5
ICOOR_INTERNAL    O7  -116.511459   66.309694    1.374530   C21   C20
C23
ICOOR_INTERNAL    H28 -179.999072   69.490695    0.994264   O7    C21
C20
```

```
ICOOR_INTERNAL    H10 -120.171312   66.625332    1.112440   C21   C20
O7
ICOOR_INTERNAL    O6  -122.097254   69.957154    1.380198   C20   C22
C21
ICOOR_INTERNAL    H27  -64.967598   73.550547    0.992744   O6    C20
C22
ICOOR_INTERNAL    H9  -124.983998   69.865294    1.119984   C20   C22
O6
ICOOR_INTERNAL    H11  120.706989   68.439637    1.115607   C22   C27
C20
ICOOR_INTERNAL    H22 -111.245097   74.994281    1.117155   C27   O8
C22
ICOOR_INTERNAL    H23 -110.732374   65.739063    1.081951   C27   O8
H22
ICOOR_INTERNAL    H26  179.382979   56.649450    1.023796   N9    C19
S2
ICOOR_INTERNAL    O2  -179.982443   59.569098    1.227206   C19   C16
N9
PDB_ROTAMERS RSB_conformers.pdb
```

FPLuc 1a:

```
NAME RSC
IO_STRING RSC Z
TYPE LIGAND
AA UNK
ATOM  C19 CNH2  X   0.81
ATOM  C16 CH1   X   0.21
ATOM  C7  aroC  X   0.08
ATOM  S1  S     X   -0.28
ATOM  C14 aroC  X   0.59
ATOM  C12 aroC  X   0.51
ATOM  N4  Nhis  X   -0.62
ATOM  C11 aroC  X   0.19
ATOM  C15 CNH2  X   0.62
ATOM  N6  Npro  X   -0.42
ATOM  C26 CH3   X   0.30
ATOM  H19 Hapo  X   0.00
ATOM  H20 Hapo  X   0.00
ATOM  H21 Hapo  X   0.00
ATOM  O1  ONH2  X   -0.57
ATOM  C18 aroC  X   -0.18
ATOM  C5  aroC  X   0.03
ATOM  C1  aroC  X   -0.15
ATOM  C3  aroC  X   -0.15
ATOM  C6  aroC  X   0.10
ATOM  C4  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  H2  Haro  X   0.15
ATOM  H4  Haro  X   0.15
ATOM  N10 Nhis  X   -0.84
ATOM  C28 CH2   X   0.37
ATOM  C24 CH3   X   0.00
ATOM  H13 Hapo  X   0.00
ATOM  H14 Hapo  X   0.00
ATOM  H15 Hapo  X   0.00
```

```
ATOM  H24 Hapo  X   0.00
ATOM  H25 Hapo  X   0.00
ATOM  C29 CH2   X   0.37
ATOM  C25 CH3   X   0.00
ATOM  H16 Hapo  X   0.00
ATOM  H17 Hapo  X   0.00
ATOM  H18 Hapo  X   0.00
ATOM  H26 Hapo  X   0.00
ATOM  H27 Hapo  X   0.00
ATOM  H3  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  H8  Haro  X   0.15
ATOM  N5  Nhis  X   -0.66
ATOM  H5  Haro  X   0.06
ATOM  N9  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O3  OOC   X   -0.65
ATOM  O4  OOC   X   -0.65
ATOM  O8  OH    X   -0.46
ATOM  C27 CH2   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C20 CH1   X   0.28
ATOM  C21 CH1   X   0.28
ATOM  C23 CH1   X   0.63
ATOM  N7  Npro  X   -0.08
ATOM  C9  aroC  X   0.55
ATOM  N3  Nhis  X   -0.66
ATOM  C10 aroC  X   0.46
ATOM  C13 aroC  X   0.78
ATOM  N2  Nhis  X   -0.66
ATOM  C8  aroC  X   0.60
ATOM  N1  Nhis  X   -0.65
ATOM  C17 CH1   X   0.51
ATOM  N8  NH2O  X   -0.12
ATOM  H28 Hpol  X   0.40
ATOM  H29 Hpol  X   0.40
ATOM  H6  Haro  X   0.06
ATOM  H7  Haro  X   0.06
ATOM  O5  OH    X   -0.56
ATOM  H12 Hapo  X   0.00
ATOM  O7  OH    X   -0.68
ATOM  H32 Hpol  X   0.40
ATOM  H10 Hapo  X   0.00
ATOM  O6  OH    X   -0.68
ATOM  H31 Hpol  X   0.40
ATOM  H9  Hapo  X   0.00
ATOM  H11 Hapo  X   0.00
ATOM  H22 Hapo  X   0.00
ATOM  H23 Hapo  X   0.00
ATOM  H30 Hpol  X   0.42
ATOM  O2  ONH2  X   -0.57
BOND_TYPE  C1   C3  4
BOND_TYPE  C1   C5  4
BOND_TYPE  C2   C4  4
BOND_TYPE  C2   C5  4
```

```
BOND_TYPE  C3   C6  4
BOND_TYPE  C4   C6  4
BOND_TYPE  C5   C18 1
BOND_TYPE  C6   N10 1
BOND_TYPE  C7   C16 1
BOND_TYPE  C7   S1  2
BOND_TYPE  C8   N1  2
BOND_TYPE  C8   N2  1
BOND_TYPE  C9   N3  1
BOND_TYPE  C9   N7  2
BOND_TYPE  C10  C13 1
BOND_TYPE  C10  C17 1
BOND_TYPE  C10  N3  2
BOND_TYPE  C11  C15 1
BOND_TYPE  C11  C18 2
BOND_TYPE  C11  N4  1
BOND_TYPE  C12  C14 1
BOND_TYPE  C12  N4  2
BOND_TYPE  C12  N6  1
BOND_TYPE  C13  N2  2
BOND_TYPE  C13  N7  1
BOND_TYPE  C14  N5  2
BOND_TYPE  C14  S1  1
BOND_TYPE  C15  N6  4
BOND_TYPE  C15  O1  2
BOND_TYPE  C16  C19 1
BOND_TYPE  C16  N5  1
BOND_TYPE  C17  N1  1
BOND_TYPE  C17  N8  1
BOND_TYPE  C19  N9  4
BOND_TYPE  C19  O2  2
BOND_TYPE  C20  C21 1
BOND_TYPE  C20  C22 1
BOND_TYPE  C20  O6  1
BOND_TYPE  C21  C23 1
BOND_TYPE  C21  O7  1
BOND_TYPE  C22  C27 1
BOND_TYPE  C22  O5  1
BOND_TYPE  C23  N7  1
BOND_TYPE  C23  O5  1
BOND_TYPE  C24  C28 1
BOND_TYPE  C25  C29 1
BOND_TYPE  C26  N6  1
BOND_TYPE  C27  O8  1
BOND_TYPE  C28  N10 1
BOND_TYPE  C29  N10 1
BOND_TYPE  N9   S2  1
BOND_TYPE  O3   S2  2
BOND_TYPE  O4   S2  2
BOND_TYPE  O8   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C4   H4  1
BOND_TYPE  C7   H5  1
```

```
BOND_TYPE  C8   H6  1
BOND_TYPE  C9   H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C20  H9  1
BOND_TYPE  C21  H10 1
BOND_TYPE  C22  H11 1
BOND_TYPE  C23  H12 1
BOND_TYPE  C24  H13 1
BOND_TYPE  C24  H14 1
BOND_TYPE  C24  H15 1
BOND_TYPE  C25  H16 1
BOND_TYPE  C25  H17 1
BOND_TYPE  C25  H18 1
BOND_TYPE  C26  H19 1
BOND_TYPE  C26  H20 1
BOND_TYPE  C26  H21 1
BOND_TYPE  C27  H22 1
BOND_TYPE  C27  H23 1
BOND_TYPE  C28  H24 1
BOND_TYPE  C28  H25 1
BOND_TYPE  C29  H26 1
BOND_TYPE  C29  H27 1
BOND_TYPE  N8   H28 1
BOND_TYPE  N8   H29 1
BOND_TYPE  N9   H30 1
BOND_TYPE  O6   H31 1
BOND_TYPE  O7   H32 1
CHI 1  C22  C20  O6   H31
PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 2  C20  C21  O7   H32
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 3  C11  C18  C5   C1
CHI 4  C3   C6   N10  C28
CHI 5  S1   C14  C12  N4
CHI 6  N9   C19  C16  C7
CHI 7  O8   C27  C22  C20
CHI 8  C21  C23  N7   C9
CHI 9  S2   O8   C27  C22
CHI 10  C6   N10  C28  C24
CHI 11  C6   N10  C29  C25
CHI 12  C19  N9   S2   O3
CHI 13  N9   S2   O8   C27
NBR_ATOM  C19
NBR_RADIUS 18.357181
ICOOR_INTERNAL    C19    0.000000    0.000000    0.000000   C19   C16
C7
ICOOR_INTERNAL    C16    0.000000  180.000000    1.496095   C19   C16
C7
ICOOR_INTERNAL    C7     0.000000   59.870629    1.372257   C16   C19
C7
ICOOR_INTERNAL    S1  -179.266651   77.664513    1.751119   C7    C16
C19
ICOOR_INTERNAL    C14    0.491524   84.357409    1.747864   S1    C7
C16
```

```
ICOOR_INTERNAL    C12  177.131963   49.951342    1.622461   C14   S1
C7
ICOOR_INTERNAL    N4     3.301075   68.968546    1.289316   C12   C14
S1
ICOOR_INTERNAL    C11  179.673582   69.437629    1.404149   N4    C12
C14
ICOOR_INTERNAL    C15    0.161299   73.173999    1.447125   C11   N4
C12
ICOOR_INTERNAL    N6     0.181994   69.783309    1.351228   C15   C11
N4
ICOOR_INTERNAL    C26  178.677200   62.669354    1.536056   N6    C15
C11
ICOOR_INTERNAL    H19  -88.595944   91.952084    1.070100   C26   N6
C15
ICOOR_INTERNAL    H20 -116.093967   68.619238    1.110427   C26   N6
H19
ICOOR_INTERNAL    H21 -128.469706   68.609586    1.110443   C26   N6
H20
ICOOR_INTERNAL    O1   179.401549   54.495891    1.223158   C15   C11
N6
ICOOR_INTERNAL    C18  179.860624   53.010262    1.348848   C11   N4
C15
ICOOR_INTERNAL    C5     0.003284   52.305008    1.495972   C18   C11
N4
ICOOR_INTERNAL    C1   120.070998   55.430092    1.406398   C5    C18
C11
ICOOR_INTERNAL    C3   179.967042   59.160770    1.397509   C1    C5
C18
ICOOR_INTERNAL    C6     0.212443   58.346730    1.413992   C3    C1
C5
ICOOR_INTERNAL    C4    -0.403339   63.071742    1.414433   C6    C3
C1
ICOOR_INTERNAL    C2     0.304625   58.736712    1.396961   C4    C6
C3
ICOOR_INTERNAL    H2  -179.781615   61.616139    1.083742   C2    C4
C6
ICOOR_INTERNAL    H4  -178.714813   57.320247    1.077185   C4    C6
C2
ICOOR_INTERNAL    N10 -179.563712   58.921400    1.451968   C6    C3
C4
ICOOR_INTERNAL    C28 -179.998648   58.921642    1.465370   N10   C6
C3
ICOOR_INTERNAL    C24  -85.908602   68.563472    1.530605   C28   N10
C6
ICOOR_INTERNAL    H13   56.324982   67.991405    1.107744   C24   C28
N10
ICOOR_INTERNAL    H14 -120.027716   70.551019    1.110234   C24   C28
H13
ICOOR_INTERNAL    H15 -119.264434   70.029409    1.110271   C24   C28
H14
ICOOR_INTERNAL    H24 -124.088393   69.562449    1.109579   C28   N10
C24
ICOOR_INTERNAL    H25 -116.943216   69.153819    1.112113   C28   N10
H24
```

```
ICOOR_INTERNAL    C29 -177.241545   57.897030    1.465675   N10   C6
C28
ICOOR_INTERNAL    C25  -81.719351   68.653456    1.528170   C29   N10
C6
ICOOR_INTERNAL    H16  -55.348410   70.204033    1.110611   C25   C29
N10
ICOOR_INTERNAL    H17 -120.954177   68.560352    1.109065   C25   C29
H16
ICOOR_INTERNAL    H18 -119.661097   70.470188    1.110361   C25   C29
H17
ICOOR_INTERNAL    H26  121.715185   66.938713    1.102697   C29   N10
C25
ICOOR_INTERNAL    H27  116.510841   70.745118    1.111779   C29   N10
H26
ICOOR_INTERNAL    H3   179.417917   63.355002    1.079223   C3    C1
C6
ICOOR_INTERNAL    H1   179.940498   57.463167    1.074843   C1    C5
C3
ICOOR_INTERNAL    H8  -179.924366   63.925149    1.087298   C18   C11
C5
ICOOR_INTERNAL    N5  -177.183569   74.691871    1.329846   C14   S1
C12
ICOOR_INTERNAL    H5   179.272402   46.637967    1.077202   C7    C16
S1
ICOOR_INTERNAL    N9    -9.928572   63.816343    1.468382   C19   C16
C7
ICOOR_INTERNAL    S2   178.603415   58.301695    1.649480   N9    C19
C16
ICOOR_INTERNAL    O3   -83.828030   72.115214    1.437728   S2    N9
C19
ICOOR_INTERNAL    O4  -110.979907   68.095120    1.445125   S2    N9
O3
ICOOR_INTERNAL    O8  -129.796485   67.884564    1.508898   S2    N9
O4
ICOOR_INTERNAL    C27   47.972804   59.847207    1.411561   O8    S2
N9
ICOOR_INTERNAL    C22  157.660576   70.738957    1.511080   C27   O8
S2
ICOOR_INTERNAL    C20 -162.092351   70.936507    1.469262   C22   C27
O8
ICOOR_INTERNAL    C21 -144.249833   77.197932    1.465575   C20   C22
C27
ICOOR_INTERNAL    C23   31.366047   77.527376    1.475916   C21   C20
C22
ICOOR_INTERNAL    N7    91.982349   65.429978    1.445332   C23   C21
C20
ICOOR_INTERNAL    C9   -93.611780   52.728629    1.352156   N7    C23
C21
ICOOR_INTERNAL    N3  -179.129471   69.965230    1.324715   C9    N7
C23
ICOOR_INTERNAL    C10    0.433861   70.203038    1.328601   N3    C9
N7
ICOOR_INTERNAL    C13   -0.110422   73.543260    1.420679   C10   N3
C9
```

```
ICOOR_INTERNAL    N2   179.726204   60.629497    1.342216   C13   C10
N3
ICOOR_INTERNAL    C8     0.608176   59.624923    1.326020   N2    C13
C10
ICOOR_INTERNAL    N1    -0.848435   57.610071    1.331554   C8    N2
C13
ICOOR_INTERNAL    C17    0.552053   58.120141    1.342096   N1    C8
N2
ICOOR_INTERNAL    N8  -179.890730   60.506917    1.449335   C17   N1
C8
ICOOR_INTERNAL    H28  179.876742   59.668192    1.030159   N8    C17
N1
ICOOR_INTERNAL    H29  179.955977   59.596310    1.030220   N8    C17
H28
ICOOR_INTERNAL    H6  -179.344486   61.198071    1.081258   C8    N2
N1
ICOOR_INTERNAL    H7   177.596100   51.693780    1.047707   C9    N7
N3
ICOOR_INTERNAL    O5  -122.223815   71.733653    1.413900   C23   C21
N7
ICOOR_INTERNAL    H12 -118.383618   70.268459    1.114938   C23   C21
O5
ICOOR_INTERNAL    O7  -116.511459   66.309694    1.374530   C21   C20
C23
ICOOR_INTERNAL    H32 -179.999072   69.490695    0.994264   O7    C21
C20
ICOOR_INTERNAL    H10 -120.171312   66.625332    1.112440   C21   C20
O7
ICOOR_INTERNAL    O6  -122.097254   69.957154    1.380198   C20   C22
C21
ICOOR_INTERNAL    H31  -64.972041   73.546813    0.992763   O6    C20
C22
ICOOR_INTERNAL    H9  -124.983998   69.865294    1.119984   C20   C22
O6
ICOOR_INTERNAL    H11  120.706989   68.439637    1.115607   C22   C27
C20
ICOOR_INTERNAL    H22 -111.245097   74.994281    1.117155   C27   O8
C22
ICOOR_INTERNAL    H23 -110.732374   65.739063    1.081951   C27   O8
H22
ICOOR_INTERNAL    H30  179.382979   56.649450    1.023796   N9    C19
S2
ICOOR_INTERNAL    O2  -179.982443   59.569098    1.227206   C19   C16
N9
PDB_ROTAMERS RSC_conformers.pdb
```

CouLuc-2-NEt_2 params file:

```
NAME LCG
IO_STRING LCG Z
TYPE LIGAND
AA UNK
ATOM  C20 CNH2  X   0.81
ATOM  C16 CH1   X   0.21
ATOM  C8  aroC  X   0.08
```

```
ATOM  S1  S     X   -0.28
ATOM  C15 aroC  X   0.58
ATOM  C19 aroC  X   -0.14
ATOM  C18 aroC  X   -0.15
ATOM  C11 aroC  X   0.01
ATOM  C7  aroC  X   -0.18
ATOM  C4  aroC  X   0.03
ATOM  C1  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C5  aroC  X   0.10
ATOM  C3  aroC  X   -0.15
ATOM  C6  aroC  X   0.08
ATOM  O5  Oaro  X   -0.23
ATOM  C14 COO   X   0.71
ATOM  O1  OOC   X   -0.57
ATOM  H3  Haro  X   0.15
ATOM  N8  Nhis  X   -0.84
ATOM  C28 CH2   X   0.37
ATOM  C25 CH3   X   0.00
ATOM  H14 Hapo  X   0.00
ATOM  H15 Hapo  X   0.00
ATOM  H16 Hapo  X   0.00
ATOM  H22 Hapo  X   0.00
ATOM  H23 Hapo  X   0.00
ATOM  C29 CH2   X   0.37
ATOM  C26 CH3   X   0.00
ATOM  H17 Hapo  X   0.00
ATOM  H18 Hapo  X   0.00
ATOM  H19 Hapo  X   0.00
ATOM  H24 Hapo  X   0.00
ATOM  H25 Hapo  X   0.00
ATOM  H2  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  H4  Haro  X   0.15
ATOM  H8  Haro  X   0.15
ATOM  H9  Haro  X   0.15
ATOM  N4  Nhis  X   -0.66
ATOM  H5  Haro  X   0.06
ATOM  N7  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O3  OOC   X   -0.65
ATOM  O4  OOC   X   -0.65
ATOM  O9  OH    X   -0.46
ATOM  C27 CH2   X   0.28
ATOM  C23 CH1   X   0.28
ATOM  C21 CH1   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C24 CH1   X   0.63
ATOM  N5  Npro  X   -0.08
ATOM  C10 aroC  X   0.55
ATOM  N3  Nhis  X   -0.66
ATOM  C12 aroC  X   0.46
ATOM  C13 aroC  X   0.78
ATOM  N2  Nhis  X   -0.66
ATOM  C9  aroC  X   0.60
```

```
ATOM  N1  Nhis  X   -0.65
ATOM  C17 CH1   X   0.51
ATOM  N6  NH2O  X   -0.12
ATOM  H26 Hpol  X   0.40
ATOM  H27 Hpol  X   0.40
ATOM  H6  Haro  X   0.06
ATOM  H7  Haro  X   0.06
ATOM  O6  OH    X   -0.56
ATOM  H13 Hapo  X   0.00
ATOM  O8  OH    X   -0.68
ATOM  H30 Hpol  X   0.40
ATOM  H11 Hapo  X   0.00
ATOM  O7  OH    X   -0.68
ATOM  H29 Hpol  X   0.40
ATOM  H10 Hapo  X   0.00
ATOM  H12 Hapo  X   0.00
ATOM  H20 Hapo  X   0.00
ATOM  H21 Hapo  X   0.00
ATOM  H28 Hpol  X   0.42
ATOM  O2  ONH2  X   -0.57
BOND_TYPE  C1   C2  4
BOND_TYPE  C1   C4  4
BOND_TYPE  C2   C5  4
BOND_TYPE  C3   C5  4
BOND_TYPE  C3   C6  4
BOND_TYPE  C4   C6  4
BOND_TYPE  C4   C7  1
BOND_TYPE  C5   N8  1
BOND_TYPE  C6   O5  1
BOND_TYPE  C7   C11 2
BOND_TYPE  C8   C16 1
BOND_TYPE  C8   S1  2
BOND_TYPE  C9   N1  2
BOND_TYPE  C9   N2  1
BOND_TYPE  C10  N3  1
BOND_TYPE  C10  N5  2
BOND_TYPE  C11  C14 1
BOND_TYPE  C11  C18 1
BOND_TYPE  C12  C13 1
BOND_TYPE  C12  C17 1
BOND_TYPE  C12  N3  2
BOND_TYPE  C13  N2  2
BOND_TYPE  C13  N5  1
BOND_TYPE  C14  O1  2
BOND_TYPE  C14  O5  1
BOND_TYPE  C15  C19 1
BOND_TYPE  C15  N4  2
BOND_TYPE  C15  S1  1
BOND_TYPE  C16  C20 1
BOND_TYPE  C16  N4  1
BOND_TYPE  C17  N1  1
BOND_TYPE  C17  N6  1
BOND_TYPE  C18  C19 2
BOND_TYPE  C20  N7  4
BOND_TYPE  C20  O2  2
```

```
BOND_TYPE  C21  C22 1
BOND_TYPE  C21  C23 1
BOND_TYPE  C21  O7  1
BOND_TYPE  C22  C24 1
BOND_TYPE  C22  O8  1
BOND_TYPE  C23  C27 1
BOND_TYPE  C23  O6  1
BOND_TYPE  C24  N5  1
BOND_TYPE  C24  O6  1
BOND_TYPE  C25  C28 1
BOND_TYPE  C26  C29 1
BOND_TYPE  C27  O9  1
BOND_TYPE  C28  N8  1
BOND_TYPE  C29  N8  1
BOND_TYPE  N7   S2  1
BOND_TYPE  O3   S2  2
BOND_TYPE  O4   S2  2
BOND_TYPE  O9   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C7   H4  1
BOND_TYPE  C8   H5  1
BOND_TYPE  C9   H6  1
BOND_TYPE  C10  H7  1
BOND_TYPE  C18  H8  1
BOND_TYPE  C19  H9  1
BOND_TYPE  C21  H10 1
BOND_TYPE  C22  H11 1
BOND_TYPE  C23  H12 1
BOND_TYPE  C24  H13 1
BOND_TYPE  C25  H14 1
BOND_TYPE  C25  H15 1
BOND_TYPE  C25  H16 1
BOND_TYPE  C26  H17 1
BOND_TYPE  C26  H18 1
BOND_TYPE  C26  H19 1
BOND_TYPE  C27  H20 1
BOND_TYPE  C27  H21 1
BOND_TYPE  C28  H22 1
BOND_TYPE  C28  H23 1
BOND_TYPE  C29  H24 1
BOND_TYPE  C29  H25 1
BOND_TYPE  N6   H26 1
BOND_TYPE  N6   H27 1
BOND_TYPE  N7   H28 1
BOND_TYPE  O7   H29 1
BOND_TYPE  O8   H30 1
CHI 1  C23  C21  O7   H29
#PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 2  C21  C22  O8   H30
#PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 3  C2   C5   N8   C28
#PROTON_CHI 3 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 4  C19  C18  C11  C7
```

```
#PROTON_CHI 4 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 5  S1   C15  C19  C18
#PROTON_CHI 5 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 6  N7   C20  C16  C8
CHI 7  O9   C27  C23  C21
CHI 8  C22  C24  N5   C10
CHI 9  S2   O9   C27  C23
CHI 10  C5   N8   C28  C25
#PROTON_CHI 10 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 11  C5   N8   C29  C26
#PROTON_CHI 11 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 12  C20  N7   S2   O3
CHI 13  N7   S2   O9   C27
NBR_ATOM  C20
NBR_RADIUS 18.224748
ICOOR_INTERNAL    C20    0.000000    0.000000    0.000000   C20   C16
C8
ICOOR_INTERNAL    C16    0.000000  180.000000    1.496021   C20   C16
C8
ICOOR_INTERNAL    C8     0.000000   59.872347    1.372318   C16   C20
C8
ICOOR_INTERNAL    S1  -179.270493   77.666388    1.751132   C8    C16
C20
ICOOR_INTERNAL    C15    0.492363   84.355877    1.747855   S1    C8
C16
ICOOR_INTERNAL    C19 -178.965358   51.788039    1.472562   C15   S1
C8
ICOOR_INTERNAL    C18   -1.305325   59.030868    1.347370   C19   C15
S1
ICOOR_INTERNAL    C11 -179.274860   53.166278    1.497084   C18   C19
C15
ICOOR_INTERNAL    C7   179.451266   63.097829    1.404762   C11   C18
C19
ICOOR_INTERNAL    C4  -179.793688   59.532570    1.395584   C7    C11
C18
ICOOR_INTERNAL    C1   179.796054   58.392558    1.400612   C4    C7
C11
ICOOR_INTERNAL    C2  -179.921545   60.040291    1.397529   C1    C4
C7
ICOOR_INTERNAL    C5    -0.116957   59.289576    1.406615   C2    C1
C4
ICOOR_INTERNAL    C3     0.075829   61.084161    1.406503   C5    C2
C1
ICOOR_INTERNAL    C6     0.052323   59.387740    1.398536   C3    C5
C2
ICOOR_INTERNAL    O5   179.736043   60.042539    1.327524   C6    C3
C5
ICOOR_INTERNAL    C14 -179.964796   56.803196    1.330734   O5    C6
C3
ICOOR_INTERNAL    O1  -179.927603   61.884600    1.215817   C14   O5
C6
ICOOR_INTERNAL    H3   178.842523   59.779401    1.082402   C3    C5
C6
ICOOR_INTERNAL    N8  -179.993686   59.454771    1.434979   C5    C2
C3
```

```
ICOOR_INTERNAL    C28   -0.002457   57.690884    1.466778   N8    C5
C2
ICOOR_INTERNAL    C25  -91.854520   64.095211    1.534414   C28   N8
C5
ICOOR_INTERNAL    H14   60.685126   69.691289    1.110806   C25   C28
N8
ICOOR_INTERNAL    H15 -118.802735   71.019176    1.110603   C25   C28
H14
ICOOR_INTERNAL    H16 -118.056018   67.538937    1.105051   C25   C28
H15
ICOOR_INTERNAL    H22 -118.146161   72.185330    1.112371   C28   N8
C25
ICOOR_INTERNAL    H23 -117.451347   70.323011    1.111477   C28   N8
H22
ICOOR_INTERNAL    C29 -179.995647   57.678678    1.466716   N8    C5
C28
ICOOR_INTERNAL    C26  -91.850411   64.105071    1.534379   C29   N8
C5
ICOOR_INTERNAL    H17   60.685925   67.549597    1.105211   C26   C29
N8
ICOOR_INTERNAL    H18 -123.147627   69.685529    1.110768   C26   C29
H17
ICOOR_INTERNAL    H19 -118.811804   71.020849    1.110601   C26   C29
H18
ICOOR_INTERNAL    H24 -118.150108   72.182516    1.112383   C29   N8
C26
ICOOR_INTERNAL    H25 -117.454404   70.317350    1.111475   C29   N8
H24
ICOOR_INTERNAL    H2  -178.850362   61.102830    1.081904   C2    C1
C5
ICOOR_INTERNAL    H1  -179.822022   59.308872    1.083577   C1    C4
C2
ICOOR_INTERNAL    H4   179.822139   59.822077    1.084739   C7    C11
C4
ICOOR_INTERNAL    H8   179.434073   62.837482    1.087641   C18   C19
C11
ICOOR_INTERNAL    H9  -179.278443   63.724142    1.078562   C19   C15
C18
ICOOR_INTERNAL    N4   178.910262   74.692820    1.329868   C15   S1
C19
ICOOR_INTERNAL    H5   179.269279   46.636101    1.077222   C8    C16
S1
ICOOR_INTERNAL    N7    -9.923225   63.813358    1.468378   C20   C16
C8
ICOOR_INTERNAL    S2   178.603099   58.300039    1.649518   N7    C20
C16
ICOOR_INTERNAL    O3   -83.829480   72.117504    1.437768   S2    N7
C20
ICOOR_INTERNAL    O4  -110.978218   68.095528    1.445070   S2    N7
O3
ICOOR_INTERNAL    O9  -129.803411   67.884839    1.508885   S2    N7
O4
ICOOR_INTERNAL    C27   47.977324   59.848471    1.411546   O9    S2
N7
```

```
ICOOR_INTERNAL    C23  157.655249   70.737770    1.511083   C27   O9
S2
ICOOR_INTERNAL    C21 -162.093482   70.933727    1.469270   C23   C27
O9
ICOOR_INTERNAL    C22 -144.247783   77.196739    1.465552   C21   C23
C27
ICOOR_INTERNAL    C24   31.367605   77.529154    1.475957   C22   C21
C23
ICOOR_INTERNAL    N5    91.983558   65.428300    1.445328   C24   C22
C21
ICOOR_INTERNAL    C10  -93.613244   52.728471    1.352163   N5    C24
C22
ICOOR_INTERNAL    N3  -179.130300   69.967367    1.324677   C10   N5
C24
ICOOR_INTERNAL    C12    0.433704   70.197497    1.328561   N3    C10
N5
ICOOR_INTERNAL    C13   -0.110640   73.547282    1.420761   C12   N3
C10
ICOOR_INTERNAL    N2   179.726171   60.632752    1.342181   C13   C12
N3
ICOOR_INTERNAL    C9     0.608583   59.621629    1.326036   N2    C13
C12
ICOOR_INTERNAL    N1    -0.848335   57.613744    1.331577   C9    N2
C13
ICOOR_INTERNAL    C17    0.553656   58.116069    1.342041   N1    C9
N2
ICOOR_INTERNAL    N6  -179.889752   60.503702    1.449271   C17   N1
C9
ICOOR_INTERNAL    H26 -152.990791   73.020370    1.019094   N6    C17
N1
ICOOR_INTERNAL    H27  125.651839   73.003358    1.011416   N6    C17
H26
ICOOR_INTERNAL    H6  -179.346779   61.198218    1.081267   C9    N2
N1
ICOOR_INTERNAL    H7   176.292279   47.119880    1.018607   C10   N5
N3
ICOOR_INTERNAL    O6  -122.227965   71.734996    1.413847   C24   C22
N5
ICOOR_INTERNAL    H13 -118.377782   70.274077    1.114966   C24   C22
O6
ICOOR_INTERNAL    O8  -116.514168   66.311108    1.374557   C22   C21
C24
ICOOR_INTERNAL    H30 -179.998770   69.489626    0.994258   O8    C22
C21
ICOOR_INTERNAL    H11 -120.121966   66.508026    1.112708   C22   C21
O8
ICOOR_INTERNAL    O7  -122.099284   69.960468    1.380236   C21   C23
C22
ICOOR_INTERNAL    H29  -64.969184   73.081754    0.992629   O7    C21
C23
ICOOR_INTERNAL    H10 -126.339895   68.634703    1.120794   C21   C23
O7
ICOOR_INTERNAL    H12  120.013706   68.521390    1.115529   C23   C27
C21
```

```
ICOOR_INTERNAL    H20 -117.558614   72.406306    1.115829   C27   O9
C23
ICOOR_INTERNAL    H21 -120.541374   68.798568    1.089626   C27   O9
H20
ICOOR_INTERNAL    H28  179.382356   56.641763    1.023812   N7    C20
S2
ICOOR_INTERNAL    O2  -179.988605   59.568862    1.227216   C20   C16
N7
```

The NapLuc-2-NMe_2 params file:

```
NAME NLA
IO_STRING NLA Z
TYPE LIGAND
AA UNK
ATOM  C21 CNH2  X   0.89
ATOM  C12 aroC  X   0.14
ATOM  C7  aroC  X   -0.11
ATOM  S1  S     X   -0.08
ATOM  C13 aroC  X   0.33
ATOM  C20 aroC  X   -0.10
ATOM  C19 aroC  X   -0.18
ATOM  C10 aroC  X   0.03
ATOM  C3  aroC  X   -0.15
ATOM  C1  aroC  X   -0.15
ATOM  C9  aroC  X   0.00
ATOM  C6  aroC  X   -0.15
ATOM  C11 aroC  X   0.10
ATOM  C4  aroC  X   -0.15
ATOM  C2  aroC  X   -0.15
ATOM  C8  aroC  X   0.00
ATOM  C5  aroC  X   -0.15
ATOM  H5  Haro  X   0.15
ATOM  H2  Haro  X   0.15
ATOM  H4  Haro  X   0.15
ATOM  N8  Nhis  X   -0.84
ATOM  C26 CH3   X   0.37
ATOM  H16 Hapo  X   0.00
ATOM  H17 Hapo  X   0.00
ATOM  H18 Hapo  X   0.00
ATOM  C27 CH3   X   0.37
ATOM  H19 Hapo  X   0.00
ATOM  H20 Hapo  X   0.00
ATOM  H21 Hapo  X   0.00
ATOM  H6  Haro  X   0.15
ATOM  H1  Haro  X   0.15
ATOM  H3  Haro  X   0.15
ATOM  H10 Haro  X   0.15
ATOM  H11 Haro  X   0.15
ATOM  N1  Nhis  X   -0.57
ATOM  H7  Haro  X   0.15
ATOM  N7  Ntrp  X   -0.79
ATOM  S2  S     X   1.62
ATOM  O2  OOC   X   -0.65
ATOM  O3  OOC   X   -0.65
```

```
ATOM  O7  OH    X   -0.46
ATOM  C28 CH2   X   0.28
ATOM  C24 CH1   X   0.28
ATOM  C22 CH1   X   0.28
ATOM  C23 CH1   X   0.28
ATOM  C25 CH1   X   0.63
ATOM  N5  Npro  X   -0.08
ATOM  C15 aroC  X   0.55
ATOM  N4  Nhis  X   -0.66
ATOM  C16 aroC  X   0.46
ATOM  C17 aroC  X   0.78
ATOM  N3  Nhis  X   -0.66
ATOM  C14 aroC  X   0.60
ATOM  N2  Nhis  X   -0.65
ATOM  C18 CH1   X   0.51
ATOM  N6  NH2O  X   -0.12
ATOM  H24 Hpol  X   0.40
ATOM  H25 Hpol  X   0.40
ATOM  H8  Haro  X   0.06
ATOM  H9  Haro  X   0.06
ATOM  O4  OH    X   -0.56
ATOM  H15 Hapo  X   0.00
ATOM  O6  OH    X   -0.68
ATOM  H28 Hpol  X   0.40
ATOM  H13 Hapo  X   0.00
ATOM  O5  OH    X   -0.68
ATOM  H27 Hpol  X   0.40
ATOM  H12 Hapo  X   0.00
ATOM  H14 Hapo  X   0.00
ATOM  H22 Hapo  X   0.00
ATOM  H23 Hapo  X   0.00
ATOM  H26 Hpol  X   0.42
ATOM  O1  ONH2  X   -0.57
BOND_TYPE  C1   C3  4
BOND_TYPE  C1   C9  4
BOND_TYPE  C2   C4  4
BOND_TYPE  C2   C8  4
BOND_TYPE  C3   C10 4
BOND_TYPE  C4   C11 4
BOND_TYPE  C5   C8  4
BOND_TYPE  C5   C10 4
BOND_TYPE  C6   C9  4
BOND_TYPE  C6   C11 4
BOND_TYPE  C7   C12 2
BOND_TYPE  C7   S1  1
BOND_TYPE  C8   C9  4
BOND_TYPE  C10  C19 1
BOND_TYPE  C11  N8  1
BOND_TYPE  C12  C21 1
BOND_TYPE  C12  N1  1
BOND_TYPE  C13  C20 1
BOND_TYPE  C13  N1  2
BOND_TYPE  C13  S1  1
BOND_TYPE  C14  N2  2
BOND_TYPE  C14  N3  1
```

```
BOND_TYPE  C15  N4  1
BOND_TYPE  C15  N5  2
BOND_TYPE  C16  C17 1
BOND_TYPE  C16  C18 1
BOND_TYPE  C16  N4  2
BOND_TYPE  C17  N3  2
BOND_TYPE  C17  N5  1
BOND_TYPE  C18  N2  1
BOND_TYPE  C18  N6  1
BOND_TYPE  C19  C20 2
BOND_TYPE  C21  N7  4
BOND_TYPE  C21  O1  2
BOND_TYPE  C22  C23 1
BOND_TYPE  C22  C24 1
BOND_TYPE  C22  O5  1
BOND_TYPE  C23  C25 1
BOND_TYPE  C23  O6  1
BOND_TYPE  C24  C28 1
BOND_TYPE  C24  O4  1
BOND_TYPE  C25  N5  1
BOND_TYPE  C25  O4  1
BOND_TYPE  C26  N8  1
BOND_TYPE  C27  N8  1
BOND_TYPE  C28  O7  1
BOND_TYPE  N7   S2  1
BOND_TYPE  O2   S2  2
BOND_TYPE  O3   S2  2
BOND_TYPE  O7   S2  1
BOND_TYPE  C1   H1  1
BOND_TYPE  C2   H2  1
BOND_TYPE  C3   H3  1
BOND_TYPE  C4   H4  1
BOND_TYPE  C5   H5  1
BOND_TYPE  C6   H6  1
BOND_TYPE  C7   H7  1
BOND_TYPE  C14  H8  1
BOND_TYPE  C15  H9  1
BOND_TYPE  C19  H10 1
BOND_TYPE  C20  H11 1
BOND_TYPE  C22  H12 1
BOND_TYPE  C23  H13 1
BOND_TYPE  C24  H14 1
BOND_TYPE  C25  H15 1
BOND_TYPE  C26  H16 1
BOND_TYPE  C26  H17 1
BOND_TYPE  C26  H18 1
BOND_TYPE  C27  H19 1
BOND_TYPE  C27  H20 1
BOND_TYPE  C27  H21 1
BOND_TYPE  C28  H22 1
BOND_TYPE  C28  H23 1
BOND_TYPE  N6   H24 1
BOND_TYPE  N6   H25 1
BOND_TYPE  N7   H26 1
BOND_TYPE  O5   H27 1
```

```
BOND_TYPE  O6   H28 1
CHI 1  C24  C22  O5   H27
#PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 2  C22  C23  O6   H28
#PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 3  C20  C19  C10  C3
#PROTON_CHI 3 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 4  C6   C11  N8   C26
#PROTON_CHI 4 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 5  N7   C21  C12  C7
CHI 6  S1   C13  C20  C19
#PROTON_CHI 6 SAMPLES 3 60 -60 180 EXTRA 1 20
CHI 7  O7   C28  C24  C22
CHI 8  C23  C25  N5   C15
CHI 9  S2   O7   C28  C24
CHI 10  C21  N7   S2   O2
CHI 11  N7   S2   O7   C28
NBR_ATOM  C21
NBR_RADIUS 16.998837
ICOOR_INTERNAL    C21    0.000000    0.000000    0.000000   C21   C12
C7
ICOOR_INTERNAL    C12    0.000000  180.000000    1.496021   C21   C12
C7
ICOOR_INTERNAL    C7     0.000000   59.872347    1.372318   C12   C21
C7
ICOOR_INTERNAL    S1  -179.270493   77.666388    1.751132   C7    C12
C21
ICOOR_INTERNAL    C13    0.492363   84.355877    1.747855   S1    C7
C12
ICOOR_INTERNAL    C20 -173.771619   61.683298    1.726380   C13   S1
C7
ICOOR_INTERNAL    C19  171.268212   57.214464    1.342253   C20   C13
S1
ICOOR_INTERNAL    C10  179.597277   55.226679    1.489884   C19   C20
C13
ICOOR_INTERNAL    C3  -160.002537   61.928409    1.404869   C10   C19
C20
ICOOR_INTERNAL    C1  -179.968581   59.087473    1.396545   C3    C10
C19
ICOOR_INTERNAL    C9    -0.006186   59.677313    1.401283   C1    C3
C10
ICOOR_INTERNAL    C6   179.987822   59.415405    1.403816   C9    C1
C3
ICOOR_INTERNAL    C11 -179.987152   58.846339    1.414986   C6    C9
C1
ICOOR_INTERNAL    C4    -0.006973   62.221683    1.413997   C11   C6
C9
ICOOR_INTERNAL    C2     0.010676   58.765899    1.396493   C4    C11
C6
ICOOR_INTERNAL    C8    -0.008063   59.529665    1.399099   C2    C4
C11
ICOOR_INTERNAL    C5   179.985804   59.390822    1.404579   C8    C2
C4
ICOOR_INTERNAL    H5     0.004226   61.554529    1.080184   C5    C8
C2
```

```
ICOOR_INTERNAL    H2  -179.996492   61.005780    1.083803   C2    C4
C8
ICOOR_INTERNAL    H4   179.988820   58.440586    1.080434   C4    C11
C2
ICOOR_INTERNAL    N8   179.998366   58.840031    1.445016   C11   C6
C4
ICOOR_INTERNAL    C26 -179.996136   59.372524    1.457096   N8    C11
C6
ICOOR_INTERNAL    H16  179.999819   67.581418    1.110235   C26   N8
C11
ICOOR_INTERNAL    H17 -119.548441   70.125007    1.110555   C26   N8
H16
ICOOR_INTERNAL    H18 -120.914948   70.131366    1.110560   C26   N8
H17
ICOOR_INTERNAL    C27  179.987400   59.307240    1.456978   N8    C11
C26
ICOOR_INTERNAL    H19  179.999960   70.119296    1.110451   C27   N8
C11
ICOOR_INTERNAL    H20 -119.532614   67.589575    1.110170   C27   N8
H19
ICOOR_INTERNAL    H21 -119.522494   70.131372    1.110635   C27   N8
H20
ICOOR_INTERNAL    H6   179.997205   61.890803    1.081476   C6    C9
C11
ICOOR_INTERNAL    H1  -179.993733   60.976718    1.083611   C1    C3
C9
ICOOR_INTERNAL    H3   179.982726   59.580568    1.083325   C3    C10
C1
ICOOR_INTERNAL    H10  179.898474   62.179240    1.084956   C19   C20
C10
ICOOR_INTERNAL    H11  179.205681   66.627727    1.084284   C20   C13
C19
ICOOR_INTERNAL    N1   173.716522   74.692820    1.329868   C13   S1
C20
ICOOR_INTERNAL    H7   179.269279   46.636101    1.077222   C7    C12
S1
ICOOR_INTERNAL    N7    -9.923225   63.813358    1.468378   C21   C12
C7
ICOOR_INTERNAL    S2   178.603099   58.300039    1.649518   N7    C21
C12
ICOOR_INTERNAL    O2   -83.829480   72.117504    1.437768   S2    N7
C21
ICOOR_INTERNAL    O3  -110.978218   68.095528    1.445070   S2    N7
O2
ICOOR_INTERNAL    O7  -129.803411   67.884839    1.508885   S2    N7
O3
ICOOR_INTERNAL    C28   47.977324   59.848471    1.411546   O7    S2
N7
ICOOR_INTERNAL    C24  157.655249   70.737770    1.511083   C28   O7
S2
ICOOR_INTERNAL    C22 -162.093482   70.933727    1.469270   C24   C28
O7
ICOOR_INTERNAL    C23 -144.247783   77.196739    1.465552   C22   C24
C28
```

```
ICOOR_INTERNAL    C25   31.367605   77.529154    1.475957   C23   C22
C24
ICOOR_INTERNAL    N5    91.983558   65.428300    1.445328   C25   C23
C22
ICOOR_INTERNAL    C15  -93.613244   52.728471    1.352163   N5    C25
C23
ICOOR_INTERNAL    N4  -179.130300   69.967367    1.324677   C15   N5
C25
ICOOR_INTERNAL    C16    0.433704   70.197497    1.328561   N4    C15
N5
ICOOR_INTERNAL    C17   -0.110640   73.547282    1.420761   C16   N4
C15
ICOOR_INTERNAL    N3   179.726171   60.632752    1.342181   C17   C16
N4
ICOOR_INTERNAL    C14    0.608583   59.621629    1.326036   N3    C17
C16
ICOOR_INTERNAL    N2    -0.848335   57.613744    1.331577   C14   N3
C17
ICOOR_INTERNAL    C18    0.553656   58.116069    1.342041   N2    C14
N3
ICOOR_INTERNAL    N6  -179.889752   60.503702    1.449271   C18   N2
C14
ICOOR_INTERNAL    H24 -152.990791   73.020370    1.019094   N6    C18
N2
ICOOR_INTERNAL    H25  125.651839   73.003358    1.011416   N6    C18
H24
ICOOR_INTERNAL    H8  -179.346779   61.198218    1.081267   C14   N3
N2
ICOOR_INTERNAL    H9   176.292279   47.119880    1.018607   C15   N5
N4
ICOOR_INTERNAL    O4  -122.227965   71.734996    1.413847   C25   C23
N5
ICOOR_INTERNAL    H15 -118.377782   70.274077    1.114966   C25   C23
O4
ICOOR_INTERNAL    O6  -116.514168   66.311108    1.374557   C23   C22
C25
ICOOR_INTERNAL    H28 -179.998770   69.489626    0.994258   O6    C23
C22
ICOOR_INTERNAL    H13 -120.121966   66.508026    1.112708   C23   C22
O6
ICOOR_INTERNAL    O5  -122.099284   69.960468    1.380236   C22   C24
C23
ICOOR_INTERNAL    H27  -64.969184   73.081754    0.992629   O5    C22
C24
ICOOR_INTERNAL    H12 -126.339895   68.634703    1.120794   C22   C24
O5
ICOOR_INTERNAL    H14  120.013706   68.521390    1.115529   C24   C28
C22
ICOOR_INTERNAL    H22 -117.558614   72.406306    1.115829   C28   O7
C24
ICOOR_INTERNAL    H23 -120.541374   68.798568    1.089626   C28   O7
H22
ICOOR_INTERNAL    H26  179.382356   56.641763    1.023812   N7    C21
S2
```

```
ICOOR_INTERNAL    O1  -179.988605   59.568862    1.227216   C21   C12
N7
```