The Role of Excited States in Determining β-lactamase Function and Bacterial Fitness

Catherine Rae Knoverek
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Biochemistry Commons, and the Biophysics Commons

Recommended Citation
Knoverek, Catherine Rae, "The Role of Excited States in Determining β-lactamase Function and Bacterial Fitness" (2021). Arts & Sciences Electronic Theses and Dissertations. 2506.
https://openscholarship.wustl.edu/art_sci_etds/2506

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
The Role of Excited States in Determining β-lactamase Function and Bacterial Fitness
by
Catherine Knoverek

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2021
St. Louis, Missouri
# Table of Contents

List of Figures ........................................................................................................... v
List of Tables ............................................................................................................... viii
List of Abbreviations ................................................................................................. ix
Acknowledgments ...................................................................................................... x
Abstract ..................................................................................................................... xv

## Chapter 1: Introduction .............................................................................................. 1

1.1 Preface ................................................................................................................... 1

1.2 Proteins are Involved in Nearly All Biological Processes .................................... 1

1.3 Proteins are Dynamic Molecules .......................................................................... 2
  1.3.1 A Brief History of Protein Dynamics ................................................................. 2
  1.3.2 Considering Protein Excited States is Important for Understanding Protein Function and Ultimately Developing Therapeutics ................................................... 6

1.4 Capturing Protein Dynamics is Notoriously Difficult ........................................... 8
  1.4.1 Current Methods for Studying Protein Dynamics .............................................. 8
  1.4.2 Remaining Questions About Protein Excited States .......................................... 10

1.5 Excited States Containing Cryptic Pockets Have Potential Utility as Drug Targets ... 12

1.6 TEM β-lactamase as a Model System .................................................................... 17

1.7 Hypotheses Addressed in this Thesis ................................................................... 19

1.8 References ............................................................................................................. 19

## Chapter 2: Opening of a Cryptic Pocket in β-lactamase Increases Penicillinase Activity ... 24

2.1 Preface ................................................................................................................... 24

2.2 Abstract .................................................................................................................. 25

2.3 Introduction ............................................................................................................ 26

2.4 Results and Discussion ........................................................................................ 29
  2.4.1 β-lactamase Homologs with Ω-Loop Pockets Hydrolyze Benzylic penicillin Faster .... 29
  2.4.2 TEM β-lactamase Variants with Higher Probabilities of Ω-Loop Pocket Opening Display Corresponding Increases in Benzylic penicillin Activity .............................. 36
2.4.3 NMR and Simulations Provide Structural Insight into the TEM Ω-Loop Pocket

Open Population Without the Need for a Mutation

2.4.4 Variants Designed to Modulate the Dynamics of the Ω-Loop Pocket Predictably

Affect Function

2.5 Conclusions

2.6 Methods

2.6.1 Mutagenesis and Protein Purification

2.6.2 Labeling Assays

2.6.3 Urea Denaturation Experiments

2.6.4 Urea Unfolding Kinetics

2.6.5 Activity Assays

2.6.6 NMR CEST Experiments

2.6.7 NMR CPMG Experiments

2.6.8 Molecular Dynamics Simulations and Analysis

2.7 References

Chapter 3: Predicting Bacterial Fitness from the Biophysical Properties of TEM β-lactamase

3.1 Abstract

3.2 Introduction

3.3 Results and Discussion

3.3.1 Catalytic Efficiency is Not Sufficient to Predict Bacterial Fitness

3.3.2 Changes in Thermodynamic Protein Stability Have Cellular Consequences

3.3.3 My Model Predicts Bacterial Fitness from Biophysical Properties

3.4 Conclusions

3.5 Methods

3.5.1 Mutagenesis and Protein Purification

3.5.2 Activity Assays

3.5.3 Urea Denaturation Experiments

3.5.4 Calculation of Fraction Folded
List of Figures

Chapter 1: Introduction

Figure 1.1 Simplified energy landscape for TEM β-lactamase.................................3

Figure 1.2 Known cryptic pockets in TEM β-lactamase.............................................13

Chapter 2: Opening of a Cryptic Pocket in β-lactamase Increases Penicillinase Activity

Figure 2.1 The Ω-loop pocket seen in TEM may open in other β-lactamase homologs.........28

Figure 2.2 β-lactamase homologs have the same topology, but only share about
50% sequence identity.................................................................30

Figure 2.3 Labeling of MTB and GNCA β-lactamases suggest that neither protein open
an Ω-loop pocket under the conditions tested.................................................31

Figure 2.4 Labeling of WT MTB is well-fit by a single exponential...............................32

Figure 2.5 Labeling of WT MTB was not due to protein unfolding...............................33

Figure 2.6 Mutating out the cysteine residue in the Ω-loop pocket region, C69, does not
reduce labeling..............................................................................33

Figure 2.7 β-lactamases with Ω-loop pockets have higher catalytic rates against
benzylpenicillin...............................................................................35

Figure 2.8 There is an inverse correlation between benzylpenicillin and cefotaxime
activity for TEM variants.....................................................................37

Figure 2.9 Simulation data was analyzed using CARDS and PCA...............................39

Figure 2.10 Structural insight into the TEM open Ω-loop pocket population identifies
conformational changes in the 238-loop and catalytic S70.................................40

Figure 2.11 The R241P mutation in TEM removes dynamics on the microsecond to
millisecond timescale as monitored by relaxation dispersion..........................41
Figure 2.12 Ω-loop pocket opening is correlated with S70 burial .................................................. 43

Figure 2.13 Mutations in TEM designed to alter the open Ω-loop pocket population lead to predictable changes in benzylpenicillin and cefotaxime activity ........................................ 45

Figure 2.14 The R241P mutation increases cefotaxime activity and decreases benzylpenicillin activity, while the E240D mutation decreases cefotaxime activity and increases benzylpenicillin activity .................................................. 46

Figure 2.15 The labeling rates for TEM R241P are on the order of magnitude as those expected for labeling due to protein unfolding, while the rates for TEM E240D are much faster ................................................................. 48

Chapter 3: Predicting Bacterial Fitness from the Biophysical Properties of TEM β-lactamase

Figure 3.1 Protein biophysics links genotype to phenotype .......................................................... 65

Figure 3.2 Catalytic efficiency accounts for 70% of the variance in minimal inhibitory concentration (MIC) ............................................................................................................. 69

Figure 3.3 Decreased stability decreases cellular protein abundance ............................................ 73

Figure 3.4 The monoclonal TEM primary antibody binds to purified TEM G238S and TEM E104K/G238S .......................................................................................................................... 74

Figure 3.5 M182T has comparable cellular protein abundance as wild-type (WT) TEM ........... 76

Figure 3.6 Considering both activity and stability predicts 99% of the variance in minimal inhibitory concentration (MIC) values .............................................................. 77

Figure 3.7 Using a threshold value of -4.3 kcal/mol explains 98% of the variance in minimal inhibitory concentration (MIC) .............................................................. 78

Chapter 4: Conclusion and Future Directions

Figure 4.1 TEM and GNCA β-lactamases share 67% sequence identity in the region of the Ω-loop pocket .................................................................................................................. 91
Figure 4.2 TEM β-lactamase has an additional cryptic pocket termed the helix 11/12 pocket..................................................................................................................................................97

Figure 4.3 Structures of cefotaxime, benzylpenicillin, and imipenem.................................103

Appendix

Figure A.1 One cysteine labels in both wild-type (WT) MTB and MTB A285C.....................108
Figure A.2 DTNB labeling WT MTB inhibits activity against nitrocefin.................................109
Figure A.3 Both cysteines label in CTX-M A283C.................................................................110
Figure A.4 The G240E mutation does not open the Ω-loop pocket in GNCA.........................110
Figure A.5 DTNB labeling of wild-type CTX-M results in a faster proteolysis rate while the presence of the compound NCI147639 results in a slower proteolysis rate...............................................................................................................................................111
Figure A.6 TEM β-lactamase folding is not reversible by thermal denaturation...................111
List of Tables

Chapter 2: Opening of a Cryptic Pocket in β-lactamase Increases Penicillinase Activity

Table 2.1 β-lactamase homologs with open Ω-loop pockets display increased catalytic rates against benzylpenicillin.................................................................34

Table 2.2 In order to determine existence of the Ω-loop pocket, the pocket population should be higher than the unfolded population.........................................................47

Chapter 3: Predicting Bacterial Fitness from the Biophysical Properties of TEM β-lactamase

Table 3.1 Catalytic efficiency (kcat/KM), thermodynamic stability (ΔGN↔I), and minimal inhibitory concentrations (MIC) for TEM variants used to determine the properties which contribute to fitness........................................72

Table 3.2 All variants are predicted to be > 99% folded.......................................................72

Table 3.3 Destabilized variants have lowered cellular protein abundance than wild-type TEM.................................................................................................73

Table 3.4 Catalytic efficiency (kcat/KM), thermodynamic stability (ΔGN↔I), and minimal inhibitory concentrations (MIC) for TEM variants used to validate the model.................................................................78
List of Abbreviations

A  alanine
a.u.  arbitrary units
C  cysteine
CARDS  correlation of all rotameric and dynamic states
CD  circular dichroism
CEST  chemical exchange saturation transfer
C_M  concentration midpoint
CPMG  Car-Purcell-Meiboom-Gill
CryoEM  cryo-electron microscopy
D  aspartic acid
DTNB  5,5'-dithiobis-(2-nitrobenzoic acid)
E  glutamic acid
FAST  fluctuation amplification of specific traits
G  glycine
K  lysine
k_2  acylation rate
k_3  deacylation rate
kcal  kilocalorie
kcat  catalytic rate
K_M  Michaelis constant
M  methionine
MSM  Markov state model
µM  micromolar
mM  millimolar
ms  millisecond
MIC  minimal inhibitory concentration
mol  mole
NMR  nuclear magnetic resonance
P  proline
PDB  Protein data bank
PBP  penicillin binding protein
PCA  principal component analysis
R  arginine
S  serine
T  threonine
Trp  tryptophan
WCL  whole cell lysate
WT  wild type
Acknowledgments

I would not have reached this point without the support and mentorship of many talented scientists who took the time to listen to, advise, and encourage me. First and foremost, I would like to thank my thesis advisor, Dr. Greg Bowman. Thank you for making me feel capable, trusted, and valued. I have appreciated your generosity with your resources and your praise. Overall, you gave me the intellectual freedom that allowed me to pursue my own ideas, and I have greatly enjoyed working, and particularly writing, with you.

I am very grateful to my thesis committee: Dr. Gautam Dantas, Dr. Michael Greenberg, Dr. Linda Pike, Dr. Christina Stallings, and Dr. Tim Wencewicz. Gautam, thank you for allowing me to be my authentic self, helping me navigate issues, teaching me how to evaluate scientific ideas, championing my science, and leading the committee skillfully. You very clearly show up for your students, and I am very lucky to have had the privilege of your mentorship. Michael, thank you for playing a pivotal role in the beginning of my graduate career. You served as my first rotation mentor and my qualifying exam committee chair, and you taught me how to develop my ideas into a thesis. Linda, you have been an inspiration and role model to me. I am grateful for your engagement and perspective in my thesis update meetings. Christina, thank you for your support and encouragement throughout the years and especially with my F31 fellowship application. Tim, thank you for all your help with my career development and ultimate job search. Your words of praise and reassurance meant a lot.
There was not a day that went by that I was not appreciative of the training I had entering into graduate school. I would like to thank Dr. Patricia Clark and Dr. Micayla Bowman for giving me so much of their time and energy when I was an undergraduate. You both were instrumental in my transformation from a student to a scientist and continue to this day to be my gold standard for what it means to be excellent in science.

Next, this thesis would not have been possible without my wonderful colleagues who are also my dear friends. Thank you, Matthew Cruz, for always being there for me. I am grateful for every conversation I have with you and every second of time I get to spend with you. You are a brilliant scientist and an overall great human. Thank you, Dr. Sukrit Singh, for the countless hours you spent with me talking through my data, interpretations, and experimental designs. You are always engaged and encouraging, and I appreciate that you were always willing to stay in lab late with me just so that I did not have to work alone. Upasana Mallimadugula, you are one of the smartest people I have ever met, and it has been a joy working with you. Thank you to Dr. Justin Miller for caring deeply about those around you and for being one of my scientific role models. I would also like to thank Dr. Thomas Frederick and Katie Moeder for being great sources of support and friendship. Thank you to everyone in the group, both past and present, who have made coming into lab truly enjoyable and without a doubt the best part of my graduate school experience.
Thank you to the students in the Biochemistry, Biophysics, and Structural Biology program for providing a collaborative, helpful, and caring community. In particular thank you to Katherine Basore and Jim Heffernan for being inspirations, for always being willing to commiserate, and most importantly for your friendship.

Finally, I would like to thank my friends and family. Thank you to all of my friends that have listened to me over the years, provided encouragement, and made sure that there has been plenty of fun. I would like to especially thank my best friend, Nora Eder. Nora, you are such an important person in my life and one of my biggest cheerleaders. Thank you so much for your friendship.

I have two wonderful parents, Louise Knoverek and Richard Knoverek, who have always made sure I have had every opportunity for success and never once doubted me. Mom, you are so intelligent, kind, and selfless. Thank you for all of the care, guidance, and support you have given to me. Dad, you are genuine, loving, and supportive. Thank you for championing my successes and always being proud of me. I would like to thank my grandma, Virginette Goodalis. You may not have always agreed with my pursuit of a graduate degree, but you taught me to be authentic and strong and that there is nothing I cannot achieve. I also remember my late stepmom, Karen Knoverek, whose love and support were always unconditional.
I am also very fortunate to have a supportive partner, Tyler Bick. Tyler, throughout my entire graduate career you have been steady, kind, and comforting. You are my best friend and my teammate, and I trust you completely. Thank you for taking care of me while also empowering me. Lastly, I would like to tell my absolutely perfect puppy, Frankie, that she is a very good girl.

Catherine Knoverek

Washington University in St. Louis

August 2021
Dedicated to my partner, Tyler Bick
ABSTRACT OF THE DISSERTATION

The Role of Excited States in Determining β-lactamase Function and Bacterial Fitness

Catherine R Knoverek

Doctor of Philosophy in Biology and Biomedical Sciences
Biochemistry, Biophysics, and Structural Biology
Washington University in St. Louis, 2021
Associate Professor Gregory R Bowman, PhD, Chair

Proteins are macromolecular machines that play a role in nearly every biological process. They are dynamic molecules which adopt many different conformations as they fold into their 3D structures, interact with their binding partners, and perform their functions. The most probable (lowest energy) protein conformation is referred to as the ground state, and this is often assumed to be the state determined by experimental methods such as x-ray crystallography. However, proteins also adopt higher energy excited states which can have significant probabilities. As these excited states are notoriously difficult to find and study, it is unclear if excited states contribute to the protein’s function and ultimately to the organism’s survival. To investigate this question, I examine the excited states of β-lactamases, enzymes which confer bacteria with antibiotic resistance by degrading β-lactam antibiotics such as penicillin. I first consider a subset of excited states that contain a cryptic pocket, or a pocket which is absent in the ground state conformation. I find that the Ω-loop cryptic pocket seen in TEM β-lactamase plays a role in the protein’s ability to hydrolyze different classes of antibiotics and specifically that the population of open pocket conformations is predictive of activity against the substrate benzylpenicillin. I next consider the subset of excited states along the protein folding pathway. I find that β-
lactamase stability contributes significantly to the ability of bacteria expressing \( \beta \)-lactamase to survive in the presence of antibiotic and specifically that including the population of folding intermediate conformations improves the predictive power of my model for bacterial fitness. These results demonstrate our understanding of how excited states are connected to \( \beta \)-lactamase function and survival of bacteria expressing \( \beta \)-lactamases. This work also helps establish the utility of cryptic pockets as drug targets and aids in our understanding of how mutations (genotype) result in changes in fitness (phenotype).
Chapter 1: Introduction

1.1 Preface

This chapter is an adapted and expanded upon version of a review published in Trends in Biochemical Sciences written by myself, Gaya Amarasinghe, and Gregory Bowman (Knoverek CR, Amarasinghe GK, Bowman GR. Advanced Methods for Accessing Protein Shape-Shifting Present New Therapeutic Opportunities. Trends Biochem Sci. 2019;44(4):351-364). As the author, I retain the right to include this work in my thesis, as outlined by the Elsevier (publisher) copyright policies.

1.2 Proteins are Involved in Nearly All Biological Processes

As a major biological macromolecule, proteins are extremely important for life. They are the molecular machines that catalyze reactions, generate force, and respond to the environment. While proteins are essential for all of the diverse biological processes, they are all made up of the same roughly 20 amino acids. Proteins have the ability to carry out so many different types of specific functions because they make use of both amino acid chemistry and a three-dimensional structure. Furthermore, proteins also sample many different structures with different likelihoods, adding even more potential for complexity in their behavior.

A strength of biochemistry research is its ability to provide molecular mechanisms for biological phenomena, and those mechanisms often involve moving proteins. Thus, understanding the ways in which a protein moves and how this relates to the protein’s function is not only a fascinating biochemical/biophysical question, but has significant implications for all biological problems. If
we can better understand the dynamics of a protein, we can better understand its function, and in turn better understand what goes wrong during disease.

1.3 Proteins are Dynamic Molecules

1.3.1 A Brief History of Protein Dynamics
A protein has the potential to adopt an enormous number of different structures. For example, a small protein with 100 amino acid residues has ~200 rotatable bonds along its backbone. Assuming that each of these rotatable bonds can adopt one of two dominant rotameric states, then such a protein can adopt approximately \(10^{60}\) different backbone structures, not to mention the additional possible structures if one considers the rotatable bonds in sidechains. Only a small fraction of these structures is capable of performing a given function. The process by which a newly synthesized or unfolded protein transitions to one of these functional structures is called protein folding and represents a particularly dramatic example of protein shape-shifting. If a protein sampled 1,000 different backbone structures each second, it would take about \(10^{40}\) times the age of the universe to enumerate every possibility. The impossibility of enumerating all possible structures is often called Levinthal’s Paradox [1]. The number of potential structures is dramatically reduced by the fact that many of the configurations considered above are infeasible because no pair of atoms can occupy the same volume in space, called the excluded volume principle. However, there is still an enormous number of different structures that do not violate the excluded volume principle.
Figure 1.1 Simplified energy landscape for TEM β-lactamase. Shown is a simplified energy landscape (gray mesh) where the y-axis represents energy, and the x-axis represents conformational space. The lowest energy structure (bottom) is the ground state, and this is often assumed to be represented by the crystal structure (PDB: 1xpb). All other structures are then referred to as excited states. Here, several known excited states for TEM β-lactamase are shown. There are two structures near the ground state that each contain an open cryptic pocket (colored orange and magenta on the structures), a structure representing the folding intermediate, and a structure representing an unfolded conformation.

The set of possible structures a protein can adopt is often conceptualized as a vast, multi-dimensional landscape, called an energy landscape (Figure 1.1). This term derives its name from the fact that each structure a protein could adopt has an associated energy that results from the interactions between different atoms in the protein and its surroundings. The probability of a protein adopting a given structure on this energy landscape is proportional to the exponential of the structure’s energy. Therefore, a protein spends exponentially more time in valleys than it does crossing the higher energy peaks separating different minima. The lowest energy structure is often referred to as the ground state and will have the highest probability of all the different...
structures a protein could adopt. Other higher-energy states are often referred to as excited states and will have lower probabilities than the ground state. Many of these excited states have negligible probabilities because their energies are so much higher than the ground state. This train of thought leaves an open question: how many different structures can a typical protein effectively adopt?

Early experiments suggested that a protein can adopt a large number of different conformations. For example, the hydrogen exchange technique (also called hydrogen-deuterium exchange, or HDX) was originally developed to test the hypothesis that amino acid residues can form helices [2]. This method leverages the fact that the bond between the nitrogen and hydrogen in the amide of the backbone of every amino acid is relatively weak. Therefore, if a protein is immersed in D$_2$O, then the hydrogens of amides that are exposed to solvent will exchange with deuterium. However, amides that form hydrogen bonds will be protected from exchange. The complex kinetics of early hydrogen exchange experiments suggested a diversity of structures were present at equilibrium.

Nuclear magnetic resonance (NMR) has also been a rich source of information on proteins’ conformational heterogeneity. NMR provides a means to assess the chemical environment of particular nuclei, often the same amide groups monitored by hydrogen exchange experiments. NMR, however, can provide information on more than just solvent exposure of the backbone amides. For example, early work showed that the phenyl groups of phenylamine residues buried in a protein’s core are capable of rotating despite the tight packing of atoms in this environment [3]. NMR can also be used for structure determination and to study other conformational
changes. However, performing these experiments is often time consuming and expensive. NMR also faces a number of technical limitations, such as the difficulty of studying large proteins.

Despite early evidence for dynamics, much of what we know about any given protein often comes from a single structure. While NMR can be used for structure determination, the first structure of a folded, globular protein was solved by x-ray crystallography [4]. In crystallography, protein is prepared under conditions that promote the formation of a regular crystal lattice. X-ray beams are then directed at the crystal, and the resulting diffraction pattern is used to infer the protein’s structure. Typically, practitioners focus on solving a single structure that best explains the diffraction data. Crystallography yields atomic resolution structures with a greater throughput than NMR, so it has come to dominate structural biology. Tens of thousands of different structures have been deposited into an online repository, called the protein data bank (PDB). These structures are often thought to represent the ground state structure in solution, but it is more accurate to think of them as the average of many low energy structures under the conditions used for crystallization. It is typically hard to capture important excited states without stabilizing the higher energy state, such as having a binding partner (i.e., small molecule or another protein). Despite this limitation, crystal structures have proved extremely valuable for gaining insight into how proteins function, as well as designing drugs and new proteins. The wealth of information a crystal structure provides sometimes even leads people to conclude that it encodes all the information one could need, rendering the role for protein dynamics negligible. For example, many methods for predicting protein stability assume that a single crystal structure is sufficient to predict the relative populations of the folded and unfolded states. Computational
drug design software also tends to make the simplifying assumption that a single protein structure is a sufficient characterization of the protein.

1.3.2 Considering Protein Excited States is Important for Understanding Protein Function and Ultimately Developing Therapeutics

Structures from techniques like crystallography and NMR are tremendously valuable, but their explanatory and predictive power is limited. The first crystal structure ever solved is a powerful illustration [4]. In this structure of myoglobin, the heme group used to bind and hold the protein’s substrate, oxygen, is buried within the protein’s core. Oxygen clearly has to get into the protein’s core to interact with heme, but it’s not obvious how it does so from this structure. Therefore, this first structure was both a demonstration of structural biology’s power and the unmet need to capture protein dynamics to achieve a complete understanding of how a protein functions. Given a single structure, it is also extremely challenging to predict essential properties for understanding a protein’s function, such as its stability [5], its affinity for different binding partners [6], or the effect of mutations on its function [7]. Often times, the crystal structures of protein variants with dramatically different activities or stabilities are essentially identical, making it difficult to explain how mutations exert their effect [8]. It is possible that this lack of explanatory and predictive power results from an inability to extract information appropriately from available structural data. However, there is mounting evidence that protein dynamics, and the conformational diversity these fluctuations give rise to, is a crucial missing factor.

If we divide the different structures a protein adopts into two classes, functional and non-functional, we can see that modulating the relative probabilities of these two classes opens a number of new therapeutic opportunities. However, this binary classification may not be adequate for many proteins, such as those with multiple functions.
Proteins that must maintain a delicate balance between populating multiple functional structures may be particularly attractive therapeutic targets because having more constraints makes them more sensitive to perturbations. For example, conformational switches involved in signaling are likely to populate at least two distinct functional states—on and off—with reasonably low energies, as well as non-functional states with higher energies. Stabilizing or destabilizing any of these states may disrupt such proteins’ ability to function appropriately. Furthermore, having more constraints to satisfy may make evolving resistance to therapeutics more challenging.

The nucleoprotein from negative sense RNA viruses presents a concrete example. For instance, Ebola virus nucleoprotein, like other negative sense RNA viral nucleoproteins, is responsible for coating the viral genome to protect it from being recognized and destroyed by a host cell. But nucleoprotein must also release RNA to allow the transcription machinery to access the viral genome. Recent work suggests that nucleoprotein can accomplish these tasks by switching between different conformations to control its affinity for RNA, and that isolated nucleoprotein has a reasonable probability of adopting both of these alternative structures in solution [9]. This balance enables nucleoprotein to serve as a context-dependent regulatory module, binding tightly to RNA until interactions with the transcription complex trigger a conformational change that favors dissociation from RNA. Furthermore, a peptide has been isolated from the transcription machinery that prevents viral replication by potently inhibiting the interaction between nucleoprotein and RNA [10]. It has been proposed that this peptide works by stabilizing nucleoprotein conformations that have a lower affinity for RNA [9]. Together, these results suggest that the relative populations of these alternative structures are constrained by the need to
switch between RNA-bound and RNA-free states and that modulating this equilibrium is a powerful therapeutic strategy. I expect many other proteins have similarly constrained equilibria and, therefore, can be targeted in a similar fashion.

1.4 Capturing Protein Dynamics is Notoriously Difficult

There is increasing consensus that trying to understand proteins without accounting for their shape-shifting is like trying to infer the rules of football from a single photo taken during a game. However, it has proven easier to acknowledge the importance of protein dynamics than to account for the role it plays in protein function. A growing community has been working to remedy this situation by developing methods to access proteins’ excited states.

1.4.1 Current Methods for Studying Protein Dynamics

Solution NMR has been particularly valuable, providing insight into both thermodynamics and kinetics. For example, relaxation dispersion NMR spectroscopy can detect the presence of an excited state with a population less than 1%. Application of this methodology to an enzyme arrested in one step of its catalytic cycle has demonstrated that the next step of the cycle is present as an excited state [11]. Such experiments have also revealed a correlation between the effects of mutations on dynamics and catalysis [11, 12]. For example, a combination of NMR, crystallography, and computer simulations have been used to map out the energy landscapes of kinases. Based on these models, it has been possible to rationalize autophosphorylation and the effects of cofactors and mutations on these enzymes’ activities [13, 14]. Initially, it was often unclear what the excited states uncovered by such experiments looked like. However, it is now possible to solve the structures of excited states [15]. Looking beyond solving the structures of
particular excited states, NMR is also being used to measure a protein’s conformational entropy as a means to quantify the number of accessible excited states [16].

Many other experimental techniques have also provided valuable insight into protein dynamics. To give a couple of examples, developments in room-temperature crystallography [17] and cryo-electron microscopy (CryoEM) [18] are providing new opportunities for obtaining high-resolution structures of excited states. Typically, crystal structures are solved based on the diffraction from a crystal at cryogenic temperatures that favor low energy structures more heavily than the temperatures where most proteins operate in vivo. Room-temperature crystallography and CryoEM both attempt to capture the distribution of structures that exists at more physiologically-relevant temperatures. Multiple structures are often required to fit the electron density detected by these methods. The relative contributions of these structures to the total density may report on their relative populations in solution. Leveraging this insight into the structures of excited states and their populations has led to improved methods for drug design [19].

Molecular dynamics simulations provide a foundation for building atomically-detailed, quantitatively predictive models of proteins’ shape-shifting that complement experiments. Such simulations are often referred to as computational microscopes because they provide a means to watch how the position of every atom in a protein evolves over time [20, 21]. The perfect simulation would provide a complete description of a protein’s thermodynamics and kinetics. However, these simulations face three key limitations: 1) the accuracy of the force fields used to model interatomic interactions, 2) the computational challenge of gathering enough data to
sample slow processes, and 3) the scientific challenge of extracting useful information from large datasets embedded in high-dimensional spaces. Significant effort has been dedicated to all of these issues and is reviewed elsewhere [22, 23].

Work from our lab, among others, has established quantitative agreement between simulations and a variety of experiments [24, 25]. These results demonstrate that existing force fields are accurate enough for many applications, given that sufficiently large datasets are collected. For example, we have shown that the agreement between different ten nanosecond simulations and NMR experiments is highly variable, but that ten microsecond simulations reliably yield quantitative agreement with these experiments [26]. In agreement with NMR and room-temperature crystallography, simulations also reveal substantial dynamics in proteins’ cores despite the tight packing that one could imagine would prevent conformational changes [27]. Due to their agreement with experiments combined with their ability to provide atomistic detail, simulations are a powerful means to explain the physical origins of experimental observations, as well as to guide the design of new experiments.

1.4.2 Remaining Questions About Protein Excited States
Even once a protein excited state has been successfully identified, it can often be difficult to determine how important that state is for the protein’s function. It is still unclear how many structures a protein typically adopts with an appreciable population and unclear how much of a protein’s energy landscape we need to know in order to have an accurate picture of the protein’s function. Because proteins are constantly evolving, it is difficult to determine which protein motions are important and which are coincidental. Furthermore, proteins interact with other
molecules, and binding events can drastically alter the protein’s energy landscape [11]. However, we currently know very little about how a protein’s energy landscape changes as it performs its function.

Critically, it is unclear if excited states generally contribute to or detract from a protein’s function. Is the ground state conformation the functional conformation and excited states provide something like stability or evolvability? Related is our poor understanding of how a protein’s dynamics are conserved or evolved over time. Either way, understanding the functional importance of excited state conformations would allow us to better identify proteins states that are useful as drug targets.

Complexity in protein biochemistry is typically added by considering multiple protein conformations or considering the protein’s biological context. However, it is still unclear how much of a protein’s *in vitro* behavior is useful for understanding its *in vivo* behavior [28]. For example, a protein may sample a given set of conformations as it folds in a test tube, but how many of those conformations are significantly populated in the cell? Cells contain a whole host of protein quality control machinery (*e.g.*, chaperones, proteases, etc.) and little is known about how excited states interact with this machinery.

These are not all of the unanswered questions related to protein dynamics. Protein energy landscapes are a relatively new concept (first studied as the protein folding problem about 60 years ago) [29], and there is still much to learn on the subject. We have just started to ask detailed questions about specific proteins of interest but are not yet ready to establish wide-
spread phenomena and guiding principles. That being said, this is an exciting time, as we know enough about protein dynamics to have some intuition and understanding and have been quickly developing new methodologies to study protein excited states. This allows us to begin to tackle some of these important questions.

1.5 Excited States Containing Cryptic Pockets Have Potential Utility as Drug Targets

The crystal structures of some proteins clearly present multiple pockets where small molecules could potentially bind and influence protein function. However, this is not always the case, and some proteins do not have any obvious druggable pockets. Fortunately, proteins’ shape-shifting can also populate excited states with pockets that are absent in available crystal structures, called cryptic pockets (Figure 1.2). These sites are called cryptic allosteric sites when they coincide with allosteric networks. Cryptic allosteric sites with coupling to key functional sites have great potential value [30]. Small molecules that bind such sites can exert allosteric control over functional sites by modulating the relative probabilities of different protein structures. They can either enhance or inhibit activity, depending on whether they increase or decrease the probability of functional conformations, respectively [31, 32].
Figure 1.2 Known cryptic pockets in TEM β-lactamase. (Left) The structure of the open helix 11/12 pocket (blue, PDB: 1pzo) is overlaid on the ligand-free crystal structure (gray, PDB: 1xpb). The region of the pocket is highlighted in orange. (Right) The structure of the open Ω-loop pocket (blue, from simulations) is overlaid on the ligand-free crystal structure (gray, PDB: 1xpb). The region of the pocket is highlighted in magenta.

Cryptic pockets can also have therapeutic value apart from allostery. For example, a cryptic pocket that forms in the middle of a key protein-protein interaction surface could be targeted to sterically block that protein-protein interaction. Cryptic pockets that extend known functional sites also provide opportunities for developing novel competitive inhibitors. For example, one could take an enzyme inhibitor that is known to bind the active site and add additional functional groups to leverage a cryptic extension to the active site.

The value of cryptic sites has not been fully realized because it is difficult to identify and target them. Most examples of cryptic pockets have been discovered serendipitously through screens that are agnostic to whether a hit binds a known functional site or a cryptic pocket. For example, a small molecule that binds in a cryptic pocket between the orange helices in Figure 1.2 was discovered through a screening campaign [33]. In this particular study, the authors sought to identify novel active site inhibitors by computationally docking a large library of chemical compounds against TEM β-lactamase’s active site and then experimentally testing the chemicals with the highest docking scores. Experimental tests of some of the top ranked compounds
confirmed that they inhibited the enzyme. However, solving the co-crystal structures of these compounds with the enzyme revealed that they did not function as intended. Instead of binding the active site, they turned out to bind a cryptic pocket.

A screening method called tethering has been developed to target a specific site on a protein, such as a cryptic site [34]. To achieve this specificity, the method requires the protein to contain a single cysteine residue near the site of interest. Satisfying this requirement often requires one or more mutations to the protein sequence. The protein is incubated with a library of chemical fragments that are capable of forming a disulfide bond with the cysteine. This disulfide tether localizes the compounds to the site of interest. Importantly, a reducing agent is also introduced along with the fragment library. This reducing agent ensures that fragments bind reversibly. Fragments that have strong non-covalent interactions with the target site will tend to stay bound even when the disulfide tether is broken by the reducing agent, while compounds that only form weak interactions with the protein will tend to dissociate. As a result, one can identify tight binding fragments by mass spectrometry. Tethering has successfully identified a number of novel compounds and pockets [35, 36]. However, a cryptic pocket could go unnoticed if the fragment library does not happen to contain any sufficiently tight binders. Moreover, it is expensive to apply tethering to multiple locations on a protein in search of a cryptic pocket. A general method for identifying cryptic sites without requiring the simultaneous discovery of compounds that bind them would be valuable for guiding the application of tethering.

Computer simulations provide an alternative approach to discover cryptic sites [37]. One of the earliest examples is the discovery of a cryptic binding trench in HIV integrase [38]. Efforts to
target this pocket eventually led to the development of raltegravir, a first-line treatment for HIV [39]. However, this success has not been replicated in a wide variety of other systems because of the challenges that molecular dynamics simulations face, as described above. A variety of techniques have been developed to overcome these limitations. Many of these methods use enhanced sampling algorithms to improve the performance of molecular dynamics simulations [40, 41]. Other approaches attempt to infer cryptic pockets from available crystal structures [42] or use alternate simulation strategies, such as the Rosetta software, to identify excited states with cryptic pockets [43].

Our lab is actively developing a pipeline that combines simulations and experiments to identify and target cryptic pockets, with an emphasis on separating the discovery of cryptic pockets from the identification of ligands that bind them. As a first step, we demonstrated that building a Markov state model (MSM) for TEM β-lactamase and applying a simple pocket detection algorithm to a representative structure for each state in the model readily identified the known cryptic pocket between the orange helices in Figure 1.2 [44]. The MSM also captured correlations between the structure of the cryptic pocket and that of the active site, consistent with the allosteric coupling between these sites. Moreover, the model predicted a multitude of new pockets with allosteric coupling to the active site. While many of these pockets are probably poor candidates for a drug design campaign [37], we proposed that a subset are potentially viable targets.

Thiol labeling experiments are a valuable means to initially test computationally predicted cryptic sites [45]. These experiments require a cysteine at a position that is buried in the apo
crystal structure but that gets exposed by the opening of a cryptic pocket. Satisfying this requirement often requires the introduction of a cysteine. However, in one case, we identified a native cysteine that satisfies these criteria, alleviating any concern that introducing a cysteine might create a cryptic pocket where none existed before [46]. Then a labeling reagent is introduced that is capable of forming a covalent bond with the cysteine if it gets exposed. An observed labeling rate that is considerably faster than that expected due to unfolding supports the existence of a cryptic pocket. For cryptic allosteric sites, one can also measure the activity of labeled protein as a first test for allosteric communication. However, the effect of labeling on activity does not necessarily determine the extent or direction of allosteric modulation that other compounds may achieve given that compounds that bind the same site can be activating, inhibiting, or have no effect on activity [31, 32].

These findings have inspired new methods to expedite the hunt for cryptic sites. For example, our lab has developed a goal-oriented adaptive sampling method, called fluctuation amplification of specific traits (FAST) [47], to identify excited states with specific geometric features more efficiently. While FAST is entirely general, one of the applications that motivated the development of the method was finding cryptic sites by searching for excited states with large pocket volumes. While our original work on cryptic pockets in TEM β-lactamase used 100 microseconds of simulation, FAST reproduces these results with just a few microseconds of simulation. Our lab has also developed a new algorithm for quickly extracting interesting excited states, such as those with cryptic pockets and cryptic allosteric sites, from large ensembles of structures generated with molecular dynamics simulations [46]. These methods have revealed yet more pockets in TEM, as well as other β-lactamases, that may be more attractive drug targets.
(Figure 1.2). In the future, we expect that incorporating quantitative measures of the druggability of cryptic sites will also be useful [37].

Assessing the conservation of cryptic sites may also be valuable. One attractive feature of sterically blocking enzyme active sites is that selective pressure to maintain function reduces the probability of mutations that are likely to disrupt inhibitor binding. If the residues lining a cryptic pocket are not constrained, then it may be easier to evolve resistance to compounds that target these sites. However, it is possible that many cryptic pockets are not so susceptible to mutation. For example, enzyme activity could be just as sensitive to mutations in cryptic allosteric sites with strong coupling to the active site as it is to mutations in the active site itself. Therefore, the conservation of residues lining different cryptic sites may also be worth considering when trying to prioritize different potential targets.

1.6 TEM β-lactamase as a Model System

TEM β-lactamase is a well-characterized enzyme which confers bacteria with antibiotic resistance by hydrolyzing, and inactivating, β-lactam antibiotics. The catalytic mechanism of this enzyme has been well-studied, including the elucidation of each individual rate constant [48]. TEM first binds substrate to form an enzyme-substrate complex. Then, the substrate is acylated and subsequently deacylated. Finally, the product is released. It is known that serine-70 is important for substrate binding and acylation, while glutamic acid-166 and a catalytic water are important for deacylation. The classical substrate of TEM is penicillins, such as benzylpenicillin, and the wild-type protein has a low catalytic efficiency against later generation antibiotics, such
as cefotaxime. However, TEM can gain increased activity against substrates like cefotaxime with only a few mutations.

Interestingly, TEM has two known cryptic pockets (Figure 1.2). As discussed above, the first pocket is located between helices 11 and 12 and was discovered serendipitously during a drug screening campaign targeting the active site [33]. The second pocket was discovered and characterized more recently by our laboratory and is located adjacent to the active site [46]. NMR [49] and room temperature crystallography [50] experiments suggest that TEM adopts multiple conformations, particularly in the region of this second cryptic pocket. Furthermore, dynamics in this region have been observed using molecular dynamics simulations and were found to be predictive of catalytic activity[7]. It is also possible that even more cryptic pockets exist in TEM, as has been predicted by Markov state models[44]. The characterization of TEM’s cryptic pockets as well as its catalytic mechanism make it an excellent model system for investigating the connection between cryptic pockets and function.

Because TEM has a singular, essential function in the cell and a selection pressure for its evolution that can be easily manipulated in the laboratory, TEM has been the focus of several studies on molecular evolution [51-53]. Moreover, between one and six mutations can change the substrate profile of the enzyme and allow it to hydrolyze new classes of antibiotics. Thus, TEM is not only a protein of interest for answering questions involving protein fitness and evolution but is also of great biomedical interest because understanding its evolution is important for combatting antibiotic resistance.
1.7 Hypotheses Addressed in this Thesis

In Chapter 2 of this thesis, I address the hypothesis that protein excited states can be beneficial for function. Specifically, I investigate excited states containing a cryptic pocket. I first test the conservation of cryptic pocket opening using thiol labeling experiments. I then gain structural insight into the pocket open population by combining NMR and molecular dynamics simulations. Finally, by designing mutations which alter the relative probabilities of the ground and excited states, I test the importance of the excited state for protein function.

In Chapter 3 of this thesis, I hypothesize that considering protein excited state conformations informs how the biophysical properties of a protein effect fitness of the organism. I determine which protein properties contribute to fitness, knowing that the properties of a protein are ultimately determined by its energy landscape. Then, I use measures of protein activity, stability, and cellular abundance in order to derive a model to accurately predict fitness.

1.8 References


Chapter 2: Opening of a Cryptic Pocket in β-lactamase Increases Penicillinase Activity

2.1 Preface

The work outlined in this chapter was a collaborative effort between me, Upasana L Mallimadugula, Sukrit Singh, Enrico Rennella, Thomas E Frederick, Tairan Yuwen, Shreya Raavicharla, Lewis E Kay, and Gregory R Bowman. I designed and coordinated the research efforts, along with LEK and GRB. I was responsible for (i) designing, producing, and purifying protein constructs, (ii) performing thiol labeling and activity assays, (iii) analyzing the data, and (iv) writing the manuscript. ULM and SR aided in the experimental design and data collection by performing stability and unfolding rate control experiments. SS performed the molecular dynamics simulations and subsequent analyses. ER, TEF, and TY performed the nuclear magnetic resonance (NMR) experiments and subsequent analyses.

This chapter has been deposited onto the bioRxiv preprint server (doi: https://doi.org/10.1101/2021.04.14.439842). As the copyright holder for this work, I authorize its reproduction here.

I would like to thank the community scientists of Folding@home for donating their computing resources for this project. This work was funded by National Institutes of Health grant R01GM12400701 (GRB), National Science Foundation CAREER Award MCB-1552471 (GRB), Canadian Institutes of Health Research FDN-503573 (LEK), and the Natural Sciences and Engineering Research Council of Canada (LEK). GRB holds a Career Award at the
Scientific Interface from the Burroughs Wellcome Fund and a Packard Fellowship for Science and Engineering from the David and Lucile Packard Foundation. LEK holds a Canada Research Chair in Biochemistry.

2.2 Abstract

Understanding the functional role of protein excited states has important implications for protein design and drug discovery. However, because these states are difficult to find and study, it is still unclear if excited states simply result from thermal fluctuations and generally detract from function or if these states can actually enhance protein function. To investigate this question, we consider excited states in β-lactamases and particularly a subset of states containing a cryptic pocket which forms under the Ω-loop. Given the known importance of the Ω-loop and the presence of this pocket in at least two homologs, we hypothesized that these excited states enhance enzyme activity. Using thiol labeling assays to probe Ω-loop pocket dynamics and kinetic assays to probe activity, we find that while this pocket is not completely conserved across β-lactamase homologs, those with the Ω-loop pocket have a higher activity against the substrate benzylpenicillin. We also find that this is true for TEM β-lactamase variants with greater open Ω-loop pocket populations. We further investigate the open population using a combination of NMR CEST experiments and molecular dynamics simulations. To test our understanding of the Ω-loop pocket’s functional role, we designed mutations to enhance/suppress pocket opening and observed that benzylpenicillin activity is proportional to the probability of pocket opening in our designed variants. The work described here suggests that excited states containing cryptic
pockets can be advantageous for function and may be favored by natural selection, increasing the potential utility of such cryptic pockets as drug targets.

2.3 Introduction

While it is well-established that proteins are dynamic molecules [1], it is often unclear what these dynamics mean for function. An experimentally derived structural snapshot of a protein, such as a crystal structure, is frequently assumed to represent the (highest probability, lowest energy) ground state. This snapshot is also frequently assumed to be the functional state of the protein. In fact, rigidifying the active site, or increasing the probability of the ground state conformation, is often used as a design strategy for improving catalytic activity [2, 3]. In opposition to this common assumption, there are several compelling examples of functionally relevant excited states [4-9]. However, it is still unclear if excited states in general play a role in function. Here, we consider an important class of excited states that contain a ‘cryptic’ pocket, or a pocket which is absent in the ligand-free, experimentally determined structure(s). These states are of particular interest because of the potential utility of cryptic pockets as drug targets [10]. These pockets provide a means to drug otherwise ‘undruggable’ proteins and a means to enhance a desired protein activity rather than just inhibit an undesired one [11, 12]. One concern, however, with the use of cryptic pockets as drug targets is that it is uncertain if there is a selective pressure to maintain the existence of a given pocket or if drug binding to that pocket could be trivially evolved away. This is at least partially because it is unknown if excited states containing cryptic pockets are simply a byproduct of the dynamic nature of proteins or if they play a bigger role in protein function.
Despite the many examples of systems which are known to contain cryptic pockets [13-15], their functional relevance remains unclear, because these pockets are notoriously difficult to find and study. Identification of a cryptic pocket often requires simultaneous discovery of a ligand that binds to it [16]. Fortunately, recent advances in computational and experimental tools allow us to better identify and study these pockets [1, 17]. To increase sampling during molecular dynamics simulations, adaptive sampling methods like FAST [18] and replica exchange methods like SWISH [19] have been developed. To analyze these datasets, methods such as Markov state models (MSMs) [20] and exposons [21] have been developed. These computational tools can then be used to inform experimental methods like room temperature crystallography [22], Nuclear Magnetic Resonance (NMR) relaxation techniques [23-25], and thiol labeling assays [26]. Previous work using these methods has shown that many different kinds of proteins have cryptic pockets and that these pockets can be targeted with drugs to allosterically affect functional sites [11, 12, 27, 28]. However, it is still unclear if cryptic pockets have implications for function in the absence of ligand binding.

To explore the functional relevance of excited states containing cryptic pockets, we consider a set of class A β-lactamases. β-lactamases are enzymes that confer bacteria with antibiotic resistance by hydrolyzing β-lactam antibiotics, such as benzylpenicillin and cefotaxime. TEM β-lactamase, in particular, is an established model system for studying cryptic pockets. TEM has two known and well-characterized cryptic pockets. The first, which was found serendipitously during a drug screening campaign, is between helices 11 and 12 [16]. The second, which was more recently identified in our laboratory [21], forms when the Ω-loop undocks from the protein,
so we call this pocket the \( \Omega \)-loop pocket (Figure 2.1). The \( \Omega \)-loop pocket was discovered in molecular dynamics simulations, confirmed using thiol labeling experiments, and subsequently shown to exert control over catalysis at the adjacent active site [21]. We know the \( \Omega \)-loop structure is important as it is necessary for the deacylation of \( \beta \)-lactam antibiotics [29], and we have previously shown that the total probability of conformations with a closed \( \Omega \)-loop is predictive of cefotaxime activity [30].

![Figure 2.1](image)

**Figure 2.1 The \( \Omega \)-loop pocket seen in TEM may open in other \( \beta \)-lactamase homologs.** The structures of four \( \beta \)-lactamase homologs (left) overlay well. TEM (PDB: 1xp) is shown in green, CTX-M-9 (PDB: 1ylj) is shown in cyan, MTB (PDB: 2gdn) is shown in orange, and GNCA (PDB: 4h88) is shown in magenta. The open \( \Omega \)-loop pocket structure in TEM (right) was identified in molecular dynamics simulations.

As we have also found that the \( \Omega \)-loop pocket is present in CTX-M-9 \( \beta \)-lactamase [21], we hypothesize that this pocket may play a role in the enzyme’s function. To test this hypothesis, we first examine if the \( \Omega \)-loop pocket is conserved across \( \beta \)-lactamase homologs and if the presence of the pocket is correlated with increased activity against classic \( \beta \)-lactam substrates. Here we mean conservation of the phenomenon of cryptic pocket opening, rather than conservation of the specific amino acid identities in that region of the protein. We then use activity data for TEM
variants and combine NMR with molecular dynamics to gain insight into how the open Ω-loop pocket affects the hydrolysis reaction for different substrates. Finally, we design mutations to modulate the population of the open Ω-loop pocket to explicitly test whether pocket dynamics are predictive of enzymatic activity.

2.4 Results and Discussion

2.4.1 β-lactamase Homologs with Ω-Loop Pockets Hydrolyze Benzylpenicillin Faster
As a first step to determining if the Ω-loop cryptic pocket seen in TEM β-lactamase is functionally relevant, we examined if this pocket is conserved across a set of β-lactamase homologs. Specifically, we examined MTB, the β-lactamase from *M. tuberculosis* encoded by the *blaC* gene, and GNCA, the predicted sequence for the last common ancestor of various Gram-negative bacteria as determined by ancestral sequence reconstruction [31]. Both of these proteins have the same topology as TEM (Figure 2.1, left), but only share about 50% sequence identity (Figure 2.2). MTB is a modern-day homolog that is evolutionarily distant from and exists in a different environment than TEM, while GNCA is a predicted ancestral sequence of TEM (Supplemental Figure 3). Investigating these two homologs allows us to test the conservation of the Ω-loop pocket in two evolutionary directions. We know that the Ω-loop pocket is conserved in at least one homolog, CTX-M-9 [21], and wider conservation would suggest there is a selective pressure to maintain the pocket because it is playing a functional role. However, even if this pocket is not perfectly conserved, we hypothesize that the presence of the pocket may be correlated with enhanced enzyme activity. While closed Ω-loop conformations of TEM β-lactamase are predictive of increased cefotaxime activity [30], there are often activity
trade-offs in enzymes [31], where increased activity for one substrate results in decreased activity for another. So, while a closed pocket may be beneficial for cefotaxime activity, an open pocket may be beneficial for a different substrate, such as benzylpenicillin.

Figure 2.2 β-lactamase homologs have the same topology, but only share about 50% sequence identity. The sequence alignment for the four homologs discussed in this study is shown here, with conserved residues shown in blue. The Ω-loop and 238-loop sequences are underlined.

To determine if the Ω-loop pocket is present in each homolog, we performed thiol labeling experiments that monitor the solvent exposure of a cysteine residue due to pocket opening. In these experiments, 5,5'-dithiobis-(2-nitrobenzoic acid), or DTNB, is added to the protein sample. DTNB covalently modifies exposed cysteine residues in a reaction that can be monitored as a change in absorbance at 412 nm over time. If a cysteine is buried inside of a pocket but then is exposed to solvent when the pocket opens, we observe an exponential increase in absorbance during the time scale of our experiment. We also ensure that the observed labeling rate is faster than the expected labeling due to protein unfolding, which is calculated from measuring the stability and/or unfolding rate. We have used this method previously to validate known [26], and identify new [21], cryptic pockets in TEM β-lactamase. As MTB and GNCA both have multiple native cysteine residues, with one cysteine located in the region of the Ω-loop pocket (Figure
2.3a-b), we added DTNB to each wild type homolog, monitored the change in absorbance, and used Beer’s law to calculate the number of cysteines labeling over time. When cysteine labeling was observed, we individually mutated out each native cysteine to determine which one is labeling in the wild type protein. For each homolog, we also measured the activity, beginning with the classic substrate, benzylpenicillin.

Figure 2.3 Labeling of MTB and GNCA β-lactamases suggest that neither protein open an Ω-loop pocket under the conditions tested. A-B. Structures of wild type (WT) MTB (PDB: 2gdn) and wild type GNCA (PDB: 4b88) are shown with native cysteine residues highlighted in gray. C69 is located in the region of the Ω-loop pocket for both proteins. C. The normalized DTNB labeling trace for WT MTB (orange circles) plateaus at one cysteine labeling. MTB C287S (light orange pentagons) shows significantly reduced labeling. D. The normalized labeling trace for WT GNCA shows no cysteine labeling.
For the conditions we tested, we find that neither MTB nor GNCA open a pocket in the region of the Ω-loop. Labeling of wild type MTB is well-fit by a single exponential (Figure 2.4), faster than expected due to unfolding (Figure 2.5), and plateaus at the expected value for one cysteine labeling (Figure 2.3c). However, when we mutate out the cysteine residue in the Ω-loop pocket region, C69, the labeling overlays well with the wild type protein (Figure 2.6). MTB C287S, however, displays significantly reduced labeling, suggesting that this is the cysteine that labels in the wild type protein. Thus, MTB does not open an Ω-loop pocket under the conditions tested here. Following the same procedure, we do not observe any significant labeling for wild type GNCA (Figure 2.3d). Again, the cysteine residue in the Ω-loop pocket region, C69, does not label, suggesting that GNCA also does not open an Ω-loop pocket under these conditions.

**Figure 2.4** Labeling of WT MTB is well-fit by a single exponential. Shown here is the average labeling trace for 30 µM protein and 2 mM DTNB. Raw data is shown as circles and the fit is shown as a solid line. Below are the residuals for the fit.
Figure 2.5 Labeling of WT MTB was not due to protein unfolding. (Left) Urea denaturation was followed by circular dichroism. Error is reported as the standard deviation of three replicate experiments. (Right) The unfolding rate was monitored by circular dichroism and then plotted as a function of urea concentration. The y-intercept of the fit line represents the unfolding rate in the absence of urea.

Figure 2.6 Mutating out the cysteine residue in the Ω-loop pocket region, C69, does not reduce labeling. The normalized DTNB labeling of MTB C69S (coral pentagons) overlays well with the labeling of WT MTB (orange circles), both which plateau at one cysteine labeling.
Table 2.1 β-lactamase homologs with open Ω-loop pockets display increased catalytic rates against benzylpenicillin.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Ω-Loop Pocket Open Population (%)</th>
<th>(\frac{k_{\text{cat}}}{K_M} (\text{sec}^{-1}\mu\text{M}^{-1}))</th>
<th>(k_{\text{cat}} (\text{sec}^{-1}))</th>
<th>(K_M (\mu\text{M}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penicillin Binding Proteins</td>
<td>N/A – no Ω-loop</td>
<td>N/A – cannot deacylate penicillin</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>GNCA β-lactamase</td>
<td>N/A – no Ω-loop pocket</td>
<td>0.40 ± 0.02</td>
<td>6.1 ± 0.1</td>
<td>15.1 ± 0.7</td>
</tr>
<tr>
<td>MTB β-lactamase</td>
<td>N/A – no Ω-loop pocket</td>
<td>0.64 ± 0.01</td>
<td>47.0 ± 0.9</td>
<td>72.9 ± 0.8</td>
</tr>
<tr>
<td>CTX-M-9 β-lactamase</td>
<td>0.023 ± 0.008‡</td>
<td>9 ± 2</td>
<td>250 ± 20</td>
<td>27 ± 4</td>
</tr>
<tr>
<td>TEM β-lactamase†</td>
<td>1.1 ± 0.2‡</td>
<td>15.2 ± 0.6</td>
<td>411 ± 5</td>
<td>27 ± 1</td>
</tr>
</tbody>
</table>

† Including an S243C mutation needed for thiol labeling measurements
‡ Porter et al., 2019

We also find, however, that the β-lactamase homologs with Ω-loop pockets display an increased ability to hydrolyze benzylpenicillin (Table 2.1 and Figure 2.7). Penicillin binding proteins, which are unable to hydrolyze benzylpenicillin, lack the Ω-loop entirely [29]. The existence of this loop, and E166 in particular, in the MTB and GNCA homologs allows for the deacylation and completed hydrolysis of benzylpenicillin with moderate catalytic efficiency. However, TEM and CTX-M-9 both not only have the Ω-loop but are able to open a pocket in this region of the protein and have a corresponding increase in their ability to hydrolyze benzylpenicillin. If the substrate binding (~\(K_M\)) but not the catalytic rate (\(k_{\text{cat}}\)) was improved, that would suggest the pocket opening simply allows for the substrate to more easily enter the active site. However, both the overall catalytic efficiency and the catalytic rate are higher for TEM and CTX-M-9 than for MTB and GNCA. As a note, we do not expect the change in magnitude of the pocket open population to perfectly correspond to the change in catalytic rate. Our thiol labeling experiments monitor the equilibrium fluctuations of the apoenzyme while several β-lactamase conformations at various ligand binding stages are sampled during our activity assays. Also, because a more open pocket is beneficial, the increase in activity is not simply due to an increase in proximity of
the catalytic residues. So, while the $\Omega$-loop pocket is not conserved across all $\beta$-lactamase homologs, the pocket appears to be functionally relevant, because benzylpenicillin hydrolysis is increased when the pocket is present.

Figure 2.7 $\beta$-lactamases with $\Omega$-loop pockets have higher catalytic rates against benzylpenicillin. The full Michaelis-Menten equation was fit to the data. Error bars are shown as the standard deviation of three replicate measurements. Error for each parameter was determined using bootstrapping.
2.4.2 TEM β-lactamase Variants with Higher Probabilities of Ω-Loop Pocket Opening Display Corresponding Increases in Benzylpenicillin Activity

To further test whether the Ω-loop pocket plays a role in the hydrolysis of benzylpenicillin, we focus our attention on a set of TEM variants. For TEM variants specifically, we have previously shown that there is a strong correlation between cefotaxime activity and the population of conformations with closed Ω-loop pockets [30]. So, we focus the rest of our study on TEM and hypothesize that an open Ω-loop pocket is detrimental for cefotaxime activity but beneficial for benzylpenicillin activity.

To begin investigating this hypothesis, we re-analyzed data from our previous study [30] where both benzylpenicillin and cefotaxime activity were measured for a number of TEM variants with mutations at clinically relevant positions. In the study, the benzylpenicillin activities were reported but not explored for their connection to the Ω-loop pocket dynamics. However, because of the correlation found between the Ω-loop closed conformations and cefotaxime activity, variants with decreased cefotaxime activity also have a higher population of conformations with an open Ω-loop pocket. Here, we examine how the catalytic rate for benzylpenicillin correlates to the catalytic efficiency against cefotaxime to determine if there is a trade-off between the two substrates, as this would suggest TEM variants with a higher population of open Ω-loop pocket conformations also have increased benzylpenicillin activity.
In fact, we do find a trade-off between benzylpenicillin and cefotaxime activity in the TEM variants examined here. We find that when TEM gains mutations that significantly increase cefotaxime activity, this corresponds to a decrease in catalytic rate for benzylpenicillin (Figure 2.8). As increased cefotaxime activity for these variants is correlated with a lower population of Ω-loop open conformations, the opposite is also true. Variants with a higher population of Ω-loop open conformations are correlated with increased benzylpenicillin activity. The variants form two distinct clusters, with the exception of TEM R164E/G238S and R164D/G238S (shown in gray), which display low activity against both substrates. Positions 164 and 238 are known to exhibit negative epistasis with one another [32], so these variants likely have perturbed Ω-loop conformations that are deleterious for all substrates.
2.4.3 NMR and Simulations Provide Structural Insight into the TEM Ω-Loop Pocket Open Population Without the Need for a Mutation

To structurally characterize the open Ω-loop pocket population, we used a combination of NMR and molecular dynamics simulations. We performed an NMR relaxation technique called Chemical Exchange Saturation Transfer (CEST) [33, 34], which allows us to observe motions on the same timescale as our labeling experiments but without the need for a cysteine mutation. CEST can identify the presence of excited states and the rates of transitioning between states, thereby complementing our thiol labeling experiments. These NMR experiments also identify which residues contribute to the exchange between the ground state and excited state(s), which we used to inform the collective variable for metadynamics simulations. We then used conformations found during our metadynamics simulations to act as starting conformations for unbiased simulations on the Folding@home distributed computing platform [28, 35]. This “adaptive seeding” strategy allows us to access protein motions on the longer timescales of our labeling and NMR experiments while still preserving the thermodynamic and kinetic properties of the system. To analyze our simulation data, we used Correlation of All Rotameric and Dynamical States (CARDS) [36] followed by Principal Component Analysis (PCA) to identify the motions associated with the Ω-loop pocket and pull out exemplar structures for the pocket open and closed populations (Figure 2.9).
Figure 2.9 Simulation data was analyzed using CARDS and PCA. (Left) Each color highlighted on the WT TEM structure (pdb: 1xpβ) represents a different CARDS community. These communities are residues which have correlated dihedral motions. The orange community includes the Ω-loop and the catalytic S70. (Right) We performed PCA on the orange community and found two resolved minima. The colored circles represent the positions of the exemplar structures shown in figure 2.10b.
Our CEST experiments provide further evidence for the TEM excited states predicted from our simulations and corroborate DTNB labeling data without the need for a mutation. Exchange was observed for TEM, suggesting the presence of an excited state population (Figure 2.10a). Many of the residues around the Ω-loop pocket report on to the exchange between the states, including L169, N170, G236, E240, G242, and S243 (Figure 2.10b, circles). The population of the excited state was determined to be $1.05 \pm 0.03\%$ ($k_{ex} = 98 \pm 6 \text{ sec}^{-1}$), which is in excellent agreement with the open pocket population measured using thiol labeling ($1.1\pm 0.2\%$). Furthermore, when
we perform the same experiment on TEM with a mutation that abolishes the Ω-loop pocket (R241P, see data below), the dynamics monitored by CEST disappear (Figure 2.10b, pentagons) as well as faster timescale dynamics monitored by relaxation dispersion (Figure 2.11). These data suggest that CEST reports on the Ω-loop pocket open population. We then used the residues identified in these experiments as being dynamic on the millisecond to second timescale to define the collective variable for metadynamics simulations, which in turn identified seed conformations for unbiased simulations.

![Figure 2.11](image)

Figure 2.11 The R241P mutation in TEM removes dynamics on the microsecond to millisecond timescale as monitored by relaxation dispersion. (Left) Highlighted in green on the wild type (WT) TEM structure (PDB: 1xpb) are the residues showing conformational exchange as established by Car-Purcell-Meiboom-Gill (CPMG) experiments. (Right) $^{15}$N CPMG profiles for wild type TEM (green circles) and TEM R241P, a variant with no Ω-loop pocket, (light green pentagons) are shown for a set of representative residues.

Analysis of our simulations show an Ω-loop pocket open state with an open Ω-loop and open 238-loop (Figure 2.10c), in good agreement with the NMR experiments reported here as well as our previous thiol labeling of this pocket [21]. We find that the Ω-loop closed conformation
looks very similar to the crystal structure, as expected. We also find that \( \Omega \)-loop pocket opening is correlated with burial of the catalytic serine (S70), making the residue no longer available to bind substrate (Figure 2.12). We observed above that closed \( \Omega \)-loop pocket conformations were generally beneficial for substrate binding, which is consistent with the closed conformation containing an S70 that is available to bind substrate. These structural results also provide insight into why a higher population of open pocket conformations may be beneficial for a larger substrate, and vice versa. Wild-type TEM \( \beta \)-lactamase degrades cefotaxime poorly and the reaction is rate-limited by the acylation rate [37, 38]. We reasoned that a more closed \( \Omega \)-loop pocket would increase protein-substrate contacts, and an available S70 would be beneficial for substrate binding and subsequent acylation. TEM hydrolyzes benzylpenicillin, on the other hand, with an efficiency approaching the diffusion limit [39, 40]. In this case, an open \( \Omega \)-loop pocket may promote product release by reducing protein-product contacts while having a minimal impact on substrate binding. Specifically, we expect a small decrease in acylation rate due to the buried S70 can be tolerated, given the already very fast acylation rate and overall small population of the pocket open conformations.
Figure 2.12 Ω-loop pocket opening is correlated with S70 burial. A 2D histogram of catalytic S70 burial as a function of Ω-loop pocket opening shows that when the pocket is open, S70 is predominantly buried. S70 burial is captured using a backbone hydrogen bond distance between S70 and K73. The Ω-loop opening distance is characterized by the Cα-Cα distance between E240 and E171.

2.4.4 Variants Designed to Modulate the Dynamics of the Ω-Loop Pocket Predictably Affect Function

Finally, we explicitly test our model by designing variants to either close or open the Ω-loop pocket, assessing the impact on pocket opening via thiol labeling experiments, and observing the effects on both benzylpenicillin and cefotaxime hydrolysis functions. We aimed to make mutations in TEM that are not known to be involved in catalysis directly but that would affect the Ω-loop pocket dynamics. Thus, changes in the pocket dynamics resulting in predictable changes in activity would support our understanding of how the TEM Ω-loop pocket is connected to function.

Towards that end, we selected Ω-loop pocket mutations based off of a previous study [41] that measured the fitness effects of every single point mutation in TEM in the presence of either
ampicillin or cefotaxime. An E240D mutation produced a positive fitness effect in the presence of ampicillin. While a mutation to lysine at this position is known to be clinically important, the charge conserving mutation to aspartic acid at this position has not been seen clinically. We rationalized that mutating this position to an aspartic acid would destabilize closed Ω-loop conformations due to its shorter hydrocarbon chain leading to lower hydrophobicity and reduced ability to screen its charged, acid group by positioning it into the solvent. Following this logic, we hypothesized that the E240D mutation opens the Ω-loop pocket in TEM, which in turn increases benzylpenicillin activity. On the other hand, a R241P mutation produced a large positive fitness effect in the presence of cefotaxime. This position is not known to be clinically important, and we reasoned that removal of a charged amino acid might stabilize closed conformations and introduction of a proline, which has fewer available dihedral angles, would destabilize open conformations. Thus, we hypothesized that the R241P mutation closes the Ω-loop pocket in TEM, which increases cefotaxime activity and decreases benzylpenicillin activity.
Figure 2.13 Mutations in TEM designed to alter the open Ω-loop pocket population lead to predictable changes in benzylpenicillin and cefotaxime activity. A. The observed labeling rate as a function of DTNB concentration is shown for wild type (WT) TEM (circles), TEM E240D (triangles), TEM R241P (pentagons), and TEM E240D/R241P (diamonds). Higher labeling rates are due to a higher open Ω-loop pocket population. The dashed line represents the expected labeling for wild type TEM due to the unfolded population. Error bars represent the standard deviation of three measurements. B. Benzylpenicillin (solid) and cefotaxime (striped) activity is shown for each variant. Error bars are the result of bootstrapping analysis.
Figure 2.14 The R241P mutation increases cefotaxime activity and decreases benzylpenicillin activity, while the E240D mutation decreases cefotaxime activity and increases benzylpenicillin activity. The full Michaelis-Menten equation was fit to the benzylpenicillin activities. A line with a slope equal to the catalytic efficiency multiplied by the enzyme concentration was fit to the cefotaxime activities. Error bars are shown as the standard deviation of three replicate measurements. Error for each parameter was determined using bootstrapping.
As predicted, we find that the E240D mutation opens the Ω-loop pocket, decreases cefotaxime activity, and increases benzylpenicillin activity (Figure 2.13 and Figure 2.14). The thiol labeling rates for TEM E240D are faster than wild type, indicating a more open Ω-loop pocket (population = 4 ± 1%, EXX regime). As expected, the catalytic efficiency of E240D for benzylpenicillin increases. The relatively small increase is also expected as TEM functions very close to the diffusion limit for this substrate. We also see the corresponding decrease in cefotaxime activity for this mutant.

Table 2.2 In order to determine existence of the Ω-loop pocket, the pocket population should be higher than the unfolded population.

<table>
<thead>
<tr>
<th>TEM Variant</th>
<th>Ω-Loop Pocket Open Population (%)</th>
<th>Unfolded Population (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild Type (S243C)</td>
<td>1.1 ± 0.2∮</td>
<td>0.02 ± 0.01∮</td>
</tr>
<tr>
<td>E240D/S423C</td>
<td>4 ± 1‡</td>
<td>0.0047 ± 0.0009</td>
</tr>
<tr>
<td>R241P/S243C</td>
<td>0.002 ± 0.005○</td>
<td>0.12 ± 0.07</td>
</tr>
<tr>
<td>E240D/R241P/S243C</td>
<td>0.19 ± 0.01</td>
<td>0.14 ± 0.02</td>
</tr>
</tbody>
</table>

∮ Porter et al., 2019
‡ Labeling in the EXX regime, individual $k_f$ and $k_r$ were determined and then divided to obtain the $K_{p口袋}$
○ Determined to not have an Ω-loop pocket as the pocket population is less than the unfolded population

We find that the R241P mutation also behaves as expected. The thiol labeling rates for TEM R241P are much slower than wild type. In fact, they are on the order of magnitude as those expected for labeling due to protein unfolding (Table 2.2 and Figure 2.15), thus suggesting that R241P abolishes the Ω-loop pocket in TEM. As a result, the activity of R241P against benzylpenicillin decreases, and we find a corresponding increase in the catalytic efficiency against cefotaxime, which is consistent with our model and suggests that the mutation does not just simply break the enzyme.
Figure 2.15 The labeling rates for TEM R241P are on the order of magnitude as those expected for labeling due to protein unfolding, while the rates for TEM E240D are much faster. Urea denaturation was followed by circular dichroism for TEM R241P (left), TEM E240D/R241P (middle), and TEM E240D (right). Error is reported as the standard deviation of three replicate experiments.

When we introduce the E240D mutation into the background of R241P, we find that the opening of the Ω-loop pocket is rescued (Figure 2.13a, population = 0.19 ± 0.01%, EX2 regime). The corresponding benzylpenicillin activity for TEM E240D/R241P is, as expected, higher than that of TEM R241P. It is also lower than that of wild type TEM, consistent with the stronger effect of the R241P mutation than that of the E240D mutation in the wild-type background. These data support our hypothesis that a higher open Ω-loop pocket population in TEM is correlated with higher benzylpenicillin activity. TEM E240D/R241P also has increased cefotaxime activity compared to wild type TEM, which is consistent with our model that closed Ω-loop conformations are beneficial for cefotaxime activity. However, the cefotaxime activity of TEM E240D/R241P, which has a low Ω-loop pocket open population, is greater than the cefotaxime activity for TEM R241P, which does not have an Ω-loop pocket. This suggests that, even though closed Ω-loop pocket conformations are beneficial for cefotaxime activity, a low open pocket population may be more advantageous than no pocket at all, or that specific closed pocket conformations additionally tune activity. Taken together, the results presented here suggest that
the Ω-loop pocket in β-lactamase plays a role in function, as excited states containing an open pocket improve activity.

2.5 Conclusions

We hypothesized that β-lactamase excited states play a role in function by increasing enzyme activity. Specifically, we investigated whether excited states containing the Ω-loop cryptic pocket enhance benzylpenicillin activity. We found that the Ω-loop pocket seen in TEM and CTX-M-9 β-lactamase is not conserved in the MTB and GNCA homologs, but that homologs with the pocket have higher catalytic rates for the hydrolysis of benzylpenicillin. Focusing our study on TEM β-lactamase, we found that variants with larger open Ω-loop pocket populations also have higher benzylpenicillin catalytic rates. To gain structural insight into the TEM open Ω-loop pocket population, we performed NMR CEST experiments and NMR-guided molecular dynamics simulations. Lastly, we designed mutations to modulate the dynamics of the Ω-loop pocket in TEM and observed that the probability of pocket opening is predictive of benzylpenicillin activity. Our results demonstrate our understanding of how excited states containing the Ω-loop cryptic pocket are connected to β-lactamase function and provide further evidence for the hypothesis that functionally relevant conformations are sampled during equilibrium fluctuations (i.e., in the apoenzyme). This work also demonstrates that cryptic pocket dynamics can be modulated with mutations, setting up future studies to elucidate the sequence determinants of these pockets, and suggests that cryptic pockets may be under positive selective pressure, increasing their potential utility as drug targets.
2.6 Methods

2.6.1 Mutagenesis and Protein Purification
We previously cloned the genes for TEM and CTX-M-9 β-lactamase into pET24-b plasmids (Life Technologies) for inducible protein expression under the T7 promoter [21, 30]. Both plasmids use kanamycin resistance for selection and the TEM plasmid contains the OmpA signal sequence for periplasmic export. MTB and GNCA were cloned into pET28 plasmids by Genewiz for inducible protein expression under the T7 promoter. Both plasmids use kanamycin resistance for selection. The MTB plasmid contains an N-terminal 6x His tag with thrombin cleavage site, and the GNCA plasmid contains the OmpA signal sequence for periplasmic export. We created protein variants using site-directed mutagenesis and verified the mutations via DNA sequencing. For expression, we transformed our desired plasmid into BL21(DE3) cells (Intact Genomics) and grew cultures to an OD$_{600}$ of 0.6 before we induced protein expression by adding 1 mM IPTG. TEM, MTB, and GNCA were expressed overnight at 18°C, while CTX-M-9 was expressed for at least three hours at 37°C.

We purified TEM and GNCA using our previously described protocol [30] that isolates the protein from the periplasm using the following osmotic shock lysis protocol. We harvested cells and resuspended them in 30 mM Tris, pH 8.0 with 20% sucrose. After centrifugation, we resuspended the cells in 5 mM MgSO$_4$ at 4°C. After another centrifugation, we dialyzed the supernatant against 20 mM sodium acetate, pH 5.5 overnight at 4°C. We centrifuged the dialysis contents to remove any insoluble protein and then purified using cation exchange chromatography (BioRad UNOsphere Rapid S column) with an NaCl gradient. The final
purification step was size exclusion chromatography (BioRad ENrich SEC 70 column), and we stored the purified protein at 4°C in 20 mM Tris, pH 8.0.

We purified CTX-M-9 as previously described [21] by isolating the protein from inclusion bodies using the following protocol. We harvested cells and resuspended them in 20 mM sodium acetate, pH 5.5 and froze them at -80°C at least overnight. We then thawed the cells and lysed them via sonication. We centrifuged the lysate, and resuspended the pellet in 20 mM sodium acetate, pH 5.5 + 9 M urea overnight. After centrifugation, we refolded the protein by adding it drop-wise to buffer with no urea while gently stirring. We removed aggregated protein by centrifugation and dialyzed the supernatant against 20 mM sodium acetate, pH 5.5 overnight at 4°C. After centrifuging the dialysis contents to remove any insoluble protein, we then purified using cation exchange chromatography (BioRad UNOsphere Rapid S column) with an NaCl gradient. The final purification step was size exclusion chromatography (BioRad ENrich SEC 70 column), and we stored the purified protein at 4°C in 20 mM Tris, pH 8.0.

We purified MTB by isolating the protein from the cytoplasm using the 6x His tag and the following protocol. Cells were harvested and resuspended in 25 mM Tris, pH 7.5 + 300 mM NaCl and frozen at -80°C overnight. Cells were then thawed and lysed via sonication. The lysate was centrifuged to remove cell debris and the supernatant was loaded onto a Ni-NTA agarose column. Elution peak fractions were then dialyzed against 25 mM Tris, pH 7.5 + 300 mM NaCl overnight at 4°C. The dialysis contents were centrifuged to remove any insoluble protein before the 6x His tag was removed by thrombin cleavage. The reaction was carried out overnight at room temperature while stirring. The reaction contents were centrifuged to remove any insoluble
protein and then cleaved protein was isolated by collecting the flow-through of a Ni-NTA agarose column run. The final purification step was size exclusion chromatography (BioRad ENrich SEC 70 column), and the purified protein was stored at 4°C in 20 mM Tris, pH 8.0.

2.6.2 Labeling Assays
For pocket determination, we labeled 5-30 µM protein with 2 mM DTNB (Ellman’s reagent, Thermo Scientific) in 20 mM Tris, pH 8.0 at 25°C until completion. We monitored the reaction via a change in absorbance at 412 nm over time using a Cary 100 UV–Vis spectrophotometer (Agilent Technologies). Each measurement was performed in triplicate. We determined the number of cysteines that labeled by first using a series of exponentials fit to the data and then by normalizing the signal using the known protein concentration and Beer’s Law (below). Here, $l$ is the pathlength of the cuvette, which is one cm, and $\varepsilon$ is the extinction coefficient of TNB at 412 nm, which is 14,150 M$^{-1}$ cm$^{-1}$ [42].

$$\text{Absorbance} = l \times \varepsilon \times [\text{protein}]$$ (2.1)

For labeling of the $\Omega$-loop pocket of the TEM variants, we labeled 10 µM protein with various concentrations of DTNB (Ellman’s reagent, Thermo Scientific) in 20 mM Tris, pH 8.0 at 25°C until completion. We introduced an S243C mutation in order to observe pocket opening via labeling, as previous described [21]. We monitored the reaction via a change in absorbance at 412 nm over time using a Cary 100 UV–Vis spectrophotometer (Agilent Technologies) and performed measurements at each DTNB concentration in triplicate. Next, we fit a single exponential equation to the data to obtain the observed rate constants and plotted these values as a function of DTNB concentration. We fit the Linderstrøm-Lang model [43] (below) to the observed rate constants as a function of DTNB concentration to obtain the open pocket
population, and obtained error using bootstrapping. We find that the TEM variants reported in this study displayed labeling in the EXX (full expression) or EX2 regime. The EX2 regime is the limiting case when the rate of pocket closing is much faster than the intrinsic rate of labeling, and the observed rate depends linearly on the DTNB concentration.

\[
k_{obs} = \frac{k_{op} \cdot k_{int} \cdot [DTNB]}{k_{op} + k_{cl} + k_{int} \cdot [DTNB]}
\]

(2.2)

\[
k_{obs,EX2} = \frac{k_{op}}{k_{cl}} \cdot k_{int} \cdot [DTNB] = K_{op/cl} \cdot k_{int} \cdot [DTNB]
\]

(2.3)

Here, \(k_{obs}\) is the observed labeling rate, \(k_{op}\) is the rate of pocket opening, \(k_{cl}\) is the rate of pocket closing, and \(k_{int}\) is the intrinsic labeling rate. We previously measured the intrinsic labeling rate (6.83 sec\(^{-1}\) mM\(^{-1}\)) by performing the same labeling experiment but on a five amino acid peptide of protein sequence containing the cysteine of interest [21]. The equilibrium constant for pocket opening (\(K_{op/cl}\)) should be greater than the equilibrium constant for unfolding, which was measured using urea denaturation experiments (see below). Using linear extrapolation, we determined the equilibrium constant for unfolding in the absence of denaturant and used the wild-type TEM unfolding equilibrium constant in equation (3) to calculate \(k_{obs}\) as a function of DTNB concentration due to unfolding (dashed line in main text Figure 2.5a). The observed rates are expected to be slow because TEM is a stable protein, meaning that the concentration of folded protein is greater than the concentration of unfolded protein and the equilibrium constant for unfolding is \(<< 1\).

### 2.6.3 Urea Denaturation Experiments

We prepared samples of 35 µg/mL protein in 50 mM potassium phosphate, pH 7.0 at various concentrations of urea and equilibrated the samples at room temperature overnight. After one-
minute incubation in an Applied Photophysics Chirascan equipped with a Quantum Northwest Inc. TC125 Peltier-controlled cuvette holder, we monitored circular dichroism (CD) signal at 222 nm. We recorded the signal for one minute at 25°C in a one cm path length quartz cuvette. Then, we measured the refractive indexes of each sample in order to determine their precise urea concentration. We determined the free energy values by fitting a two-state folding model (below) to the CD data and using the linear extrapolation method [44]. Each variant was measured in triplicate experiments.

\[
CD = \frac{\theta_u + \theta_f \ast e^{-(\Delta G + M \ast \text{[urea]}) / R \ast T}}{1 + e^{-(\Delta G + M \ast \text{[urea]}) / R \ast T}} \tag{2.9}
\]

Here, \(\theta_u\) and \(\theta_f\) are the CD signals for the unfolded and folded states, fit as lines. \(\Delta G\) is the extrapolated free energy difference between the unfolded and folded states in the absence of urea, and \(M\) is the proportionality constant related to the steepness of the folding transition. \(R\) is the gas constant and \(T\) is temperature.

### 2.6.4 Urea Unfolding Kinetics
We prepared a protein sample in 50 mM potassium phosphate, pH 7.0 and samples of various urea concentrations above the concentration midpoint (\(C_M\)) in the same buffer. After five-minute incubation in an Applied Photophysics Chirascan equipped with a Quantum Northwest Inc. TC125 Peltier-controlled cuvette holder, we added the protein to the urea buffer, diluting the protein to a final concentration of 35 \(\mu\)g/mL, and manually mixed by inverting the cuvette. We then monitored the circular dichroism (CD) signal at 222 nm over time at 25°C in a one cm path length quartz cuvette. We measured the refractive indexes of each sample in order to determine their precise urea concentration. An exponential was fit to each unfolding kinetic trace to determine the observed unfolding rate at that given urea concentration. The unfolding rate in
the absence of urea was then calculated using a linear extrapolation fit to the log of the observed unfolding rates as a function of urea.

2.6.5 Activity Assays
We measured the initial velocity ($v_i$) of antibiotic degradation by $\beta$-lactamase at 25°C via a change in absorbance (232 nm for benzylpenicillin, 262 nm for cefotaxime) using a Cary 100 UV–Vis spectrophotometer (Agilent Technologies). The substrate (5-200 µM) was incubated at 25°C for 5 min before addition of the protein. We diluted purified protein to a final concentration no greater than 200 nM. Our activity buffer was 50 mM potassium phosphate, pH 7.0 with 10% glycerol, and we measured each substrate concentration in triplicate. For benzylpenicillin, the Michaelis-Menten equation (below) was fit to the initial velocity as a function of the substrate concentration to determine individual catalytic rate ($k_{cat}$) and Michaelis constant ($K_M$) values. Here, $[E]$ is the total enzyme concentration, and $[S]$ is the total substrate concentration. For cefotaxime, the $K_M$ was too high to reach maximum velocity, so a line (below) with a slope equal to $[E] \times \frac{k_{cat}}{K_M}$ was fit to the data. Error for the fit parameters was determined using bootstrapping.

\[
v_i = \frac{[E] \times k_{cat} \times [S]}{K_M + [S]} \quad (2.4)
\]

\[
v_i = [E] \times \frac{k_{cat}}{K_M} \times [S] \quad (2.5)
\]

2.6.6 NMR CEST Experiments
We recorded all experiments on a Bruker AVANCE III HD 18.8 T spectrometer equipped with a cryogenically cooled, x,y,z pulsed-field gradient triple-resonance probe. We recorded each D-CEST [45, 46] experiment (30°C) as a pseudo-3D matrix, where each 2D spectrum were
obtained as a function of the position of weak B\textsubscript{1} perturbations applied at discrete frequencies over the chemical shift range of the probed nucleus. We applied a DANTE excitation scheme [47], which perturbs multiple regularly spaced frequencies at the same time, thereby decreasing the frequency range that must be explored over regular CEST approaches. In all cases, we calibrated the strength of the B\textsubscript{1} field using a nutation experiment, as described previously [48].

We acquired \textsuperscript{15}N D-CEST data as previously described [45] using 1 s DANTE excitation trains of square pulses (~7° flip angle, 2.5 kHz B\textsubscript{1} field) and an interpulse delay of 2, 1, and 0.667 ms, resulting in effective B\textsubscript{1} fields of about 10, 20, and 30 Hz. We sampled CEST profiles in 51 steps, with increments of 10, 20, and 30 Hz, extending over frequency ranges of 500 (2 ms interpulse delay), 1000 (1 ms), and 1500 Hz (0.667 ms), respectively. We extracted the position of the minor dips, exchange rate, and population of the excited state by fitting a two-state model of chemical exchange to the CEST data as described in detail previously [46]. Errors were estimated using the bootstrapping.

2.6.7 NMR CPMG Experiments
We recorded \textsuperscript{15}N CPMG experiments as previously described [49], using a constant-time relaxation interval, T\textsubscript{relax}, of 30 milliseconds [50]. We sampled 20 v\textsubscript{CPMG} values of \leq 1 kHz, using CPMG refocusing pulses applied at a $\gamma B_1/2\pi = 6$ kHz field and phase-modulated according to the \{x,x,y,-y\} cycling scheme [51]. We applied a $\gamma B_1/2\pi = 15.6$ kHz field \textsuperscript{1}H continuous wave decoupling during T\textsubscript{relax}. To ensure constant heating in the reference experiments (recorded with T\textsubscript{relax} equal to 0 seconds), we applied the same \textsuperscript{1}H continuous wave decoupling immediately prior to the recycle delay.
2.6.8 Molecular Dynamics Simulations and Analysis
We prepared systems as previously described [20], using GROMACS software [52] and the Amber03 force field [53]. We solvated in TIP3P water [54] and energy minimized using the steepest descent algorithm. We used the V-rescale thermostat to maintain a fixed temperature of 300K and the Berendsen barostat to bring the pressure up to one bar. Mutations were introduced into the starting structure using PyMol.

We ran 200 nanoseconds of metadynamics simulations on each variant using the PLUMED plugin on GROMACS [55] and defined our collective variable (s) using the backbone torsional angles of the 238-loop. This collective variable is expressed using the equation below.

\[ s = \frac{1}{2} \sum_{i=1}^{n} [1 + \cos(\phi_i - \phi_i^{Ref})] \] (2.6)

Here, \( \phi_i \) is the torsion angle of dihedral \( i \) of the current frame in the simulation and \( \phi_i^{Ref} \) is a reference torsion angle defined by the corresponding dihedral in the crystal structure. This summation is done across \( n \) dihedrals. We computed our collective variable using the \( \phi \) and \( \psi \) angles of residues that underwent CEST, namely residues 237 through 243. Gaussians were added every 2 picoseconds with a height of 1.0 kJ/mol and a width of 0.05.

We clustered our metadynamics simulations using a hybrid k-centers/k-medoids algorithm [56] to generate 220 representative seed conformations, using a cluster radius cutoff of 1.2 Å. We then used these conformations to collect a total of 100.7 microseconds of unbiased simulations on our Folding@home distributed computing platform [57].
To analyze our simulation data, we first built a Markov state model (MSM) using our Enspara software [58]. We again clustered using a hybrid k-centers/k-medoids algorithm and used a cluster radius cutoff of 1.2 Å, which resulted in 9877 states. A pseudo-count was added to each element in the transition counts matrix to prevent sampling artifacts from influencing the transition probabilities.

We then used the CARDS methodology [36] to compute the holistic communication \( I_H(X, Y) \) for every pair of dihedrals \( X \) and \( Y \) using the equation below.

\[
I_H(X, Y) = I_{ss}(X,Y) + I_{sd}(X,Y) + I_{ds}(X,Y) + I_{dd}(X,Y)
\]  
(2.7)

Here, \( I_{ss}(X,Y) \) is the normalized mutual information between the structure (i.e., rotameric state) of dihedral \( X \) and the structure of dihedral \( Y \), \( I_{sd}(X,Y) \) is the normalized mutual information between the structure of dihedral \( X \) and the dynamical state of dihedral \( Y \), \( I_{ds}(X,Y) \) is the normalized mutual information between the dynamical state of dihedral \( X \) and the structure of dihedral \( Y \), and \( I_{dd}(X,Y) \) is the normalized mutual information between the dynamical state of dihedral \( X \) and the dynamical state of dihedral \( Y \). The mutual information \( I \) is described by the equation below.

\[
I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right)
\]  
(2.8)

Here, \( x \in X \) refers to the set of possible states that dihedral \( X \) can adopt, \( p(x) \) is the probability that dihedral \( X \) adopts state \( x \), and \( p(x,y) \) is the joint probability that dihedral \( X \) adopts state \( x \) and dihedral \( Y \) adopts state \( y \). We computed normalized mutual information using the maximum possible mutual information, known as the channel capacity, for any specific mode of communication. We then computed a community network using affinity propagation [59], with a
damping parameter of 0.8. We generated the final allosteric network by filtering the community network using the Marginal Likelihood Filter (MLF) [60] to capture the top 5% of edges.

Finally, we applied principal component analysis (PCA) to the distances between the Cβ atoms of every pair of residues in the community containing the Ω-loop and catalytic S70. We projected our MSM onto principal components 1 and 3 (PC1 and PC3) and pulled out exemplar structures by estimating the population-weighted centroid of the two minima.

### 2.7 References


Chapter 3: Predicting Bacterial Fitness from the Biophysical Properties of TEM β-lactamase

3.1 Abstract

One strategy for closing the genotype-phenotype gap is to consider protein biophysics, as mutational effects can often be observed in measurable protein properties, and functional proteins are essential for organism fitness. However, it is still uncertain which protein properties contribute to fitness and to what magnitude. Here, I hypothesize that due to the multiple conformations a protein adopts and its complex biological context, both protein activity and stability are needed to predict fitness. To address my hypothesis, I investigated how the properties of TEM β-lactamase, a bacterial enzyme responsible for degrading β-lactam antibiotics, affect bacterial fitness as measured by minimal inhibitory concentration (MIC). I used a set of clinically prevalent TEM variants that increase bacteria resistance to the third-generation cephalosporin cefotaxime. I found that protein activity alone, measured as catalytic efficiency, only explains 70% of the variance in MIC. By measuring TEM protein stability using three different signals, I found that TEM has a folding intermediate which is susceptible to proteolytic digestion. I reasoned that changes in the concentration of this intermediate lead to the changes seen in cellular protein abundance. As describing protein stability using fraction folded is insufficient to predict the abundance differences between variants, I derived an expression that describes the concentration of folded, active enzyme given the existence and proteolytic
susceptibility of a folding intermediate. Ultimately, I found that my model explains 99% of the variance in MIC.

### 3.2 Introduction

Disease often occurs because a mutation in a gene encoding a protein alters or disrupts that protein’s function. However, it is often unclear which mutations cause disease or the molecular mechanism in which the mutations exert their effects. In the simplest case, a mutation first changes the biophysical properties (e.g., activity and stability) of a singular protein and then those changes affect how well the protein functions and the organism survives. As mutations typically affect more than one biophysical property (pleiotropy) and their effects are dependent on the background protein sequence (epistasis) [1], it is difficult to predict the biophysical effects of a mutation. Furthermore, it is often unclear which proteins and which properties of the proteins significantly contribute to the fitness of the organism [2]. This problem in general is referred to as the genotype-phenotype gap (Figure 3.1).

**Figure 3.1. Protein biophysics links genotype to phenotype.** Shown is a visual representation of how genotype and phenotype are related. A mutation is a change in genotype that can lead to a change in phenotype such as organism fitness. As mutations can affect the properties of a protein and how well a protein functions can be important for fitness, the biophysical properties of a protein can help bridge the genotype-phenotype gap.
As proteins link genotype and phenotype and have in vitro measurable properties, several of the previous successes towards closing the genotype-phenotype gap have focused on the biophysical properties of a protein essential for fitness [2-4]. However, proteins exist in often complicated in vivo contexts, and it is still unclear which in vitro properties of a protein can be used to predict fitness. In particular, it is unclear if protein stability contributes to fitness and how useful in vitro measurements of stability are in predicting fitness [5]. As proteins are dynamic molecules [6], I hypothesize that considering multiple protein conformations, along with the protein’s biological context, is necessary to close the genotype-phenotype gap.

To address this hypothesis for one protein of interest, I consider how the biophysical properties of TEM β-lactamase lead to changes in E.coli fitness. As a naturally plasmid-encoded enzyme with a clear selective pressure, TEM β-lactamase is widely used to study molecular evolution [5, 7-12]. TEM is a periplasmic enzyme that degrades and inactivates β-lactam antibiotics, such as penicillin. These antibiotics bind to penicillin binding proteins (PBPs), the proteins responsible for building and maintaining the cell wall. Without TEM, the β-lactam antibiotics inhibit PBPs and the cells burst [13]. Thus, TEM β-lactamase plays a singular, essential role in the cell.

Here, I use the mutations E104K, M182T, and G238S, alone and in combination. These mutations are known to significantly increase bacterial resistance to the third-generation cephalosporin, cefotaxime, are three of the most clinically prevalent mutations, and have been seen clinically in all combinations [5]. As my measure of bacteria fitness, I use minimal inhibitory concentration (MIC) of cefotaxime. MIC is a discrete value and has been shown to be correlated with other measures of fitness, such as growth rate [11]. I then use the biophysical
properties of the TEM variants to predict fitness by determining the minimal set of properties that contribute to fitness, deriving a model for predicting fitness using those properties, and then testing my model with a set of new variants.

3.3 Results and Discussion

3.3.1 Catalytic Efficiency is Not Sufficient to Predict Bacterial Fitness
I aimed to determine the minimal number of protein biophysical properties that are necessary to predict bacterial fitness. Accurate prediction would demonstrate our understanding of the protein properties that contribute to fitness and that ultimately describe the molecular context for mutations. As β-lactamase function is essential for survival in the presence of antibiotics, I began with the hypothesis that TEM β-lactamase catalytic efficiency \( (k_{cat}/K_M) \) alone is sufficient to predict MIC.

Nikaido and Normark in 1987 proposed, and subsequently showed, that MIC values could be accurately predicted by combining the expression for the diffusion of the antibiotic across the outer membrane (Fick diffusion) with the expression describing the hydrolysis of the antibiotic by β-lactamase enzymes (Michaelis-Menten) [14]. Their model is given below.

\[
MIC = [S]_{lethal} + \frac{[S]_{lethal} \times k_{cat} \times [N]}{P \times A \times (K_M + [S]_{lethal})}
\]

(3.1)

Here \([S]_{lethal}\) is the concentration of antibiotic needed to kill the cells (based on antibiotic binding to essential PBPs) [14], and \([N]\) is the concentration of natively folded protein. \(k_{cat}\) and \(K_M\) are the catalytic rate and Michaelis constant, respectively, determined from Michaelis-Menten kinetics. \(P\) is the outer membrane permeability to the antibiotic, and \(A\) is the surface area.
of the cell. Nikaido and Normark predicted MIC by measuring the $v_{max}$ (equal to $k_{cat} \times [N]$) of β-lactamases from cellular extracts [14]. However, if we follow Knies, Cai, and Weinreich and assume that $[N]$ is constant because TEM and its variants are stable (>99% folded) and that $K_M \gg [S]_{lethal}$, the model can be simplified to the one shown below [5].

$$MIC = Z \times \frac{k_{cat}}{K_M}$$  \hspace{1cm} (3.2)

$Z$ includes $[S]_{lethal}$, $[N]$, $P$, and $A$, which are all assumed to be independent on the TEM variant being produced, and thus my expression for MIC depends only on properties of TEM that can be measured in vitro.

To test my hypothesis, I produced and purified each TEM variant from *E. coli* to measure the catalytic efficiency against cefotaxime using Michaelis-Menten kinetics (Table 3.1). I then measured the minimal inhibitory concentration of bacteria producing each TEM variant in the presence of cefotaxime under the native TEM promoter and signal sequence (Table 3.1). Finally, I predicted the MIC values from the catalytic efficiencies by fitting the single $Z$ parameter, which I found to be $31 \pm 6 \text{ µM^2 sec}$. 

68
Catalytic efficiency accounts for 70% of the variance in minimal inhibitory concentration (MIC). Catalytic efficiency was measured for each TEM variant using Michaelis-Menten kinetics. Using a single parameter to account for TEM-independent factors, MIC was predicted (y-axis) and correlated to experimental MIC (x-axis). The dashed line represents a perfect 1:1 correlation.

I found that catalytic efficiency alone predicts only 70% of the variance in MIC (Figure 3.2). As expected, there is a strong correlation ($R^2 = 0.7 \pm 0.1$) between catalytic efficiency and MIC. However, I also found that a large increase in fitness could be achieved without a corresponding increase in activity. TEM E104K/G238S has a catalytic efficiency of $3.0 \pm 0.1$ sec$^{-1}$µM$^{-1}$ and a MIC of 36 µM, while TEM E104K/M182T/G238S has a similar catalytic efficiency of $3.13 \pm 0.09$ sec$^{-1}$µM$^{-1}$ but a much higher MIC of 144 µM. Thus, my results suggest that another property significantly contributes to bacterial fitness. This result is not surprising given that TEM mutations are known to display epistasis (non-additivity) in their fitness effects, leading to a rough fitness landscape [5, 7, 15].
3.3.2 Changes in Thermodynamic Protein Stability Have Cellular Consequences

As beneficial mutations typically increase fitness by either increasing protein activity or increasing the concentration of properly folded protein in the cell, and cellular protein abundance is modified by thermodynamic stability [3], I hypothesized that TEM stability significantly contributes to bacterial fitness. In agreement with this hypothesis, previous work by Jacquier et al. found that 29% of the variance in MIC could be predicted by TEM stability alone, suggesting that combining TEM catalytic efficiency and stability should theoretically account for 99% of the variance in MIC. Wild-type TEM, however, is very stable (in vitro global thermodynamic stability of -14.3 ± 0.3 kcal/mol) [5, 16], and it has been previously shown that measurements of thermostability do not aid in the prediction of MIC [5]. That being said, stability values are likely different in vivo than in vitro because different protein conformations have different likelihoods of aggregating, degrading, and interacting with other cellular components. I hypothesized that changes in stability have physiological consequences that can be modeled using in vitro stability measurements.

I first measured TEM stability of my purified protein using three independent signals: (i) circular dichroism (CD) reporting on α-helical secondary structure, (ii) intrinsic tryptophan fluorescence reporting on the change in the environment of aromatic residues, and (iii) gel band intensity after pulse proteolysis reporting on proteolytic conformations. I used urea as a chemical denaturant in all three experiments, determined free energy differences using a linear extrapolation model (see Methods), and compared each measurement using fraction folded.
As shown previously, I found that TEM does not fold in a two-state manner (Figure 3.3a). An intermediate population can be observed from the CD experiments. When monitored via fluorescence, unfolding appears two-state, but transitions at a urea concentration near the first transition seen by CD. In fact, I have previously globally fit to CD and fluorescence data using a three-state model to extract stability values for each transition. Furthermore, when I monitor stability using pulse proteolysis experiments, I find that the folding transition overlays well with the fluorescence data. These data suggest that TEM folds via at least a three-state mechanism. Fluorescence and pulse proteolysis report on the same transition from the native to the intermediate populations, suggesting that the intermediate population is proteolytically susceptible.

As I found that fluorescence reports on the first (lowest energy) visible folding transition, I measured the stability using this method for each of the TEM variants (Table 3.1). Strikingly, I found that the two variants highlighted above, TEM E104K/G238S and TEM E104K/M182T/G238S, have very different stabilities (Figure 3.3b). This might explain their differences in MIC values, as TEM E104K/M182T/G238S has both a larger stability and a larger MIC. However, given their free energy values, both proteins are predicted to be > 99% folded (Table 3.2).
Table 3.1. Catalytic efficiency ($k_{\text{cat}}/K_m$), thermodynamic stability ($\Delta G_{N\leftrightarrow I}$), and minimal inhibitory concentrations (MIC) for TEM variants used to determine the properties which contribute to fitness.

<table>
<thead>
<tr>
<th>Variant</th>
<th>$k_{\text{cat}}/K_m$ for Cefotaxime ($\mu$M$^{-1}$ sec$^{-1}$)</th>
<th>$\Delta G_{N\leftrightarrow I}$ (kcal/mol)</th>
<th>MIC (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild-type TEM</td>
<td>0.0020 ± 0.0005$^*$</td>
<td>-6.0 ± 0.2</td>
<td>0.28</td>
</tr>
<tr>
<td>TEM E104K</td>
<td>0.012 ± 0.004$^*$</td>
<td>-5.7 ± 0.4</td>
<td>0.28</td>
</tr>
<tr>
<td>TEM M182T</td>
<td>0.0018 ± 0.0002$^*$</td>
<td>-10.0 ± 0.7</td>
<td>0.28</td>
</tr>
<tr>
<td>TEM G238S</td>
<td>0.29 ± 0.03$^*$</td>
<td>-3.6 ± 0.2</td>
<td>9</td>
</tr>
<tr>
<td>TEM E104K/M182T</td>
<td>0.0108 ± 0.0003</td>
<td>-8.8 ± 0.2</td>
<td>0.625</td>
</tr>
<tr>
<td>TEM E104K/G238S</td>
<td>3.0 ± 0.1</td>
<td>-3.0 ± 0.3</td>
<td>36</td>
</tr>
<tr>
<td>TEM M182T/G238S</td>
<td>0.22 ± 0.03</td>
<td>-6.1 ± 0.5</td>
<td>9</td>
</tr>
<tr>
<td>TEM E104K/M182T/G238S</td>
<td>3.13 ± 0.09</td>
<td>-5.4 ± 0.2</td>
<td>144</td>
</tr>
</tbody>
</table>

$^*$Hart et al., 2016

To test whether a stability change of about 3 kcal/mol can alter cellular protein abundance, I produced TEM variants in *E. coli* under the native TEM promotor and signal sequence. Then I lysed the cells and separated proteins using gel electrophoresis. Gel samples were normalized by total protein concentration as measured via a Bradford assay. I visualized only the bands corresponding to TEM by performing a Western blot with TEM primary antibodies. As a test case, I measured the cellular protein abundance of wild-type TEM and the two variants that displayed the lowest *in vitro* stabilities, TEM G238S ($\Delta G_{N\leftrightarrow I} = -3.6 \pm 0.2$ kcal/mol) and TEM E104K/G238S ($\Delta G_{N\leftrightarrow I} = -3.0 \pm 0.3$ kcal/mol).

Table 3.2. All variants are predicted to be > 99% folded.

<table>
<thead>
<tr>
<th>Variant</th>
<th>$\Delta G_{N\leftrightarrow I}$ (kcal/mol)</th>
<th>Fraction Folded</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild-type TEM</td>
<td>-6.0 ± 0.2</td>
<td>0.9999</td>
</tr>
<tr>
<td>TEM E104K</td>
<td>-5.7 ± 0.4</td>
<td>0.9999</td>
</tr>
<tr>
<td>TEM M182T</td>
<td>-10.0 ± 0.7</td>
<td>0.9999</td>
</tr>
<tr>
<td>TEM G238S</td>
<td>-3.6 ± 0.2</td>
<td>0.9977</td>
</tr>
<tr>
<td>TEM E104K/M182T</td>
<td>-8.8 ± 0.2</td>
<td>0.9999</td>
</tr>
<tr>
<td>TEM E104K/G238S</td>
<td>-3.0 ± 0.3</td>
<td>0.9937</td>
</tr>
<tr>
<td>TEM M182T/G238S</td>
<td>-6.1 ± 0.5</td>
<td>0.9999</td>
</tr>
<tr>
<td>TEM E104K/M182T/G238S</td>
<td>-5.4 ± 0.2</td>
<td>0.9998</td>
</tr>
</tbody>
</table>
Figure 3.3. Decreased stability decreases cellular protein abundance. A) Wild-type TEM folding/unfolding was monitored using circular dichroism (blue circles), intrinsic tryptophan (Trp) fluorescence (orange triangles), and band intensity after pulse proteolysis (gray pentagons). Fraction folded is shown as a function of urea concentration for each representative experiment. Free energy values were determined from three independent experiments and fit using a linear extrapolation model. Fits are shown as solid lines in corresponding colors. B) Stability was measured via fluorescence for TEM E104K/G238S (blue circles) and TEM E104K/M182T/G238S (orange triangles). Fraction folded is shown as a function of urea concentration for each representative experiment. Free energy values were determined from three independent experiments and fit using a linear extrapolation model. Fits are shown as solid lines in corresponding colors. C) Cellular protein abundance was measured for whole cell lysate (WCL) and soluble fractions using a Western blot with anti-TEM antibodies. The molecular weights of the pre-stained protein ladder are shown on the left. Bands can be seen at the expected molecular weight for wild-type TEM (WT TEM) but not for TEM G238S (G238S) or TEM E104K/G238S (E104K/G238S). D) Shown is the kinetic scheme used to derive my model describing minimal inhibitory concentration. U, I, and N are the unfolded, intermediate, and native states, respectively. Their rates of folding and unfolding are described using the same convention. $k_{syn}$ is the synthesis rate, and $k_{prot}$ is the proteolysis rate.

Table 3.3. Destabilized variants have lowered cellular protein abundance than wild-type TEM.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Relative Cellular Protein Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild-type TEM</td>
<td>100</td>
</tr>
<tr>
<td>TEM G238S</td>
<td>0.78 ± 0.08</td>
</tr>
<tr>
<td>TEM E104K/G238S</td>
<td>0.0040 ± 0.0002</td>
</tr>
</tbody>
</table>

I found that my destabilized variants displayed drastically reduced cellular abundance (Figure 3.3c). Clear bands were visible for wild-type TEM from both the whole cell lysate and soluble...
samples. However, no bands were visible for TEM G238S or TEM E104K/G238S despite confirmed antibody binding (Figure 3.4). This is likely due to protein cellular abundance levels below the detection limit of my experiment as both of these variants have higher MIC values than wild-type TEM and lowered abundance was quantified using an independent assay (Table 3.3). Decreased thermodynamic stability likely decreased cellular protein abundance by either causing the protein to aggregate or be degraded. As the whole cell lysate and soluble samples display similar intensities for each variant, and I showed above that the intermediate population is susceptible to proteolytic digestion (Figure 3.3a), I reasoned that TEM variants with decreased stabilities displayed lowered cellular abundance due to their larger populations of the intermediate being degraded by proteases. This appears to be true despite the percent of intermediate population in vitro calculated to be < 1%.

![Figure 3.4](image.png)

**Figure 3.4. The monoclonal TEM primary antibody binds to purified TEM G238S and TEM E104K/G238S.** While no antibody binding is detected with cell lysate samples after protein expression, the antibody is shown to bind to purified protein for TEM G238S (left) and TEM E104K/G238S (right).
3.3.3 My Model Predicts Bacterial Fitness from Biophysical Properties

As I showed above that protein stability affects protein abundance in the cell, but that differences in fraction folded between my variants are very small, I derived a different expression for stability that could be used in my model. I first allowed \([N]\) to depend on the TEM variant produced and considered a three-state folding mechanism where TEM is first synthesized to an unfolded state, U, folds to an intermediate state, I, and then proceeds to a native state, N (Figure 3.3c). Besides the folding \((k_{U\rightarrow I} \text{ and } k_{I\rightarrow N})\) and unfolding rates \((k_{I\rightarrow U} \text{ and } k_{N\rightarrow I})\), TEM can also be synthesized with a rate \(k_{\text{syn}}\) and degraded from the intermediate state with a rate \(k_{\text{prot}}\). I assumed the concentration of U to be negligible, that \(k_{\text{syn}}\) and \(k_{\text{prot}}\) are independent of the TEM variant produced, and that \([N]\) reaches steady-state. This resulted in the expression for MIC given below.

\[
MIC = Z' \cdot \frac{k_{\text{cat}}}{K_M} \cdot e^{-\Delta G_{N\rightarrow I}/RT} \tag{3.3}
\]

\(\Delta G_{N\rightarrow I}\) is the free energy difference between the native and intermediate populations, \(R\) is the gas constant, and \(T\) is temperature.
**Figure 3.5. M182T has comparable cellular protein abundance as wild-type (WT) TEM.** Cellular protein abundance was measured using a Western blot with anti-TEM antibodies. The molecular weights of the pre-stained protein ladder are shown on the left. Similar intensity bands can be seen at the expected molecular weight for wild-type TEM (WT TEM) and TEM M182T (M182T). The bands are also comparable between the whole cell lysate (WCL) and soluble samples.

I then predicted MIC using the measured values in Table 1 and by fitting the $Z'$ parameter. I also fit an additional parameter which describes a stability threshold. Practically, this parameter results from the assumption that increased stability past a certain threshold does not increase fitness. This is based on the observation that the M182T mutation has comparable fitness (Table 3.1) and cellular abundance (Figure 3.5) to wild-type TEM despite increasing stability by about 4 kcal/mol. Stabilizing mutations have also been shown in general to be neutral alone but beneficial when they compensate for activating, destabilizing mutations [10]. Mathematically, this parameter results from inclusion of an exportation rate of the native protein from the periplasm.
I found that my model predicted 99% of the variance (R² = 0.9999 ± 0.0003) in MIC (Figure 3.6, blue circles). Strikingly, consideration of stability resulted in both TEM E104K/G238S and TEM E104K/M182T/G238S to fall on the 1:1 correlation line. In the case of TEM beta-lactamase, I found that activity accounts for 70% of the variance in MIC and stability accounts for 29%, in agreement with previous work [11]. The value of my Z’ parameter was 0.078 ± 0.001 and my threshold was -3.781 ± 0.009 kcal/mol. If I consider that wild-type TEM may actually only have 1.7 kcal/mol stability in vivo [11], this would mean that my in vitro stability measurements are over-estimating stability by about 4.3 kcal/mol. My fit threshold was similar to this value and using a -4.3 kcal/mol threshold yields similar results (Figure 3.7).
Figure 3.7. Using a threshold value of -4.3 kcal/mol explains 98% of the variance in minimal inhibitory concentration (MIC). MIC was predicted (y-axis) and correlated to experimental MIC (x-axis) using my model that considers both protein activity and stability. I used the previously fit (main text) $Z^\prime$ parameter and a value of -4.3 kcal/mol as my threshold value. The dashed line represents a perfect 1:1 correlation. Error in the correlation was determined using bootstrapping.

Encouraged by these results, I next aimed to validate my model with a set of new TEM variants. I was interested to see if my model would be generalizable to non-clinical mutations, so I selected the A42G mutation, alone and in combination with my other single mutants. A42G, in contrast to my other mutations, is not known to be clinically important, was instead identified during an in vitro evolution experiment, and has also been much less studied. This mutation appears to be relatively neutral in the wild-type background, but displayed positive epistasis when in combination with the other three mutations studied here.

Table 3.4. Catalytic efficiency ($k_{cat}/K_m$), thermodynamic stability ($\Delta G_{N\rightarrow I}$), and minimal inhibitory concentrations (MIC) for TEM variants used to validate the model.

<table>
<thead>
<tr>
<th>Variant</th>
<th>$k_{cat}/K_m$ for Cefotaxime ((\mu\text{M}^{-1} \text{sec}^{-1}))</th>
<th>$\Delta G_{N\rightarrow I}$ (kcal/mol)</th>
<th>MIC ((\mu\text{M}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEM A42G</td>
<td>0.00276 ± 0.00004</td>
<td>-6.2 ± 0.6</td>
<td>0.14</td>
</tr>
<tr>
<td>TEM A42G/E104K</td>
<td>0.0368 ± 0.0003</td>
<td>-6.1 ± 0.2</td>
<td>0.56</td>
</tr>
<tr>
<td>TEM A42G/M182T</td>
<td>0.00286 ± 0.00006</td>
<td>-9.4 ± 0.4</td>
<td>0.14</td>
</tr>
<tr>
<td>TEM A42G/G238S</td>
<td>0.58 ± 0.04</td>
<td>-4.5 ± 0.3</td>
<td>10</td>
</tr>
</tbody>
</table>
Given the activity and stability measurements of the new TEM variants (Table 3.2) and using the same values for $Z'$ and the threshold that were fit above, I found that my model was able to accurately predict MIC values (Figure 3.6, orange triangles). Specifically, I found the correlation for all variants to be $R^2 = 0.99 \pm 0.04$. In summary, my results show that both protein activity and stability are necessary for predicting bacterial fitness, suggesting that these properties contribute significantly to fitness and describe the molecular context for mutations in this system.

### 3.4 Conclusions

I aimed to predict *E. coli* bacterial fitness using the biophysical properties of TEM β-lactamase. I found that catalytic efficiency alone was not sufficient to explain the variance in MIC, particularly because a large gain in MIC could be achieved without a corresponding change in catalytic efficiency. I next considered the role of protein stability in fitness. I measured stability using three independent signals and found that a $< 3$ kcal/mol destabilization between the native and proteolytically susceptible intermediate populations *in vitro* was enough to drastically reduce protein abundance *in vivo*. This finding provides evidence for the importance of higher energy protein states in determining cellular protein concentrations and ultimately organism fitness. Finally, I derived a model for predicting MIC using *in vitro* activity and stability measurements of TEM and found that I could predict 99% of the variance in MIC. The exact values of my fit parameters are likely to be improved upon as more TEM variant data is collected. However, overall, my results demonstrate my understanding of the protein biophysical properties which contribute to fitness, and I expect my model to be generalizable to other β-lactam antibiotics.
3.5 Methods

3.5.1 Mutagenesis and Protein Purification
I previously cloned the gene for TEM-1 β-lactamase into a pET24-b plasmid (Life Technologies) with kanamycin resistance and OmpA signal sequence for inducible protein expression under the T7 promoter [17]. I generated protein variants using site-directed mutagenesis and verified the mutations via DNA sequencing. I then produced and purified TEM variants as previously described [17]. Briefly, I transformed my desired plasmid into BL21(DE3) cells (Intact Genomics), grew cultures to an OD$_{600}$ of 0.6, and induced protein production with 1 mM IPTG. I expressed overnight at 18°C before purifying TEM from the periplasm using osmotic shock. I harvested cells and equilibrated them first in 30 mM Tris, pH 8.0 with 20% sucrose at room temperature and then in 5 mM MgSO$_4$ at 4°C. I next dialyzed the supernatant against 20 mM sodium acetate, pH 5.5 overnight at 4°C in preparation for purification the next day using cation exchange chromatography (BioRad UNOsphere Rapid S column) with an NaCl gradient. Finally, I performed size exclusion chromatography (BioRad ENrich SEC 70 column) and stored the purified protein at 4°C in 20 mM Tris, pH 8.0.

3.5.2 Activity Assays
Catalytic efficiency was measured as described previously [18]. Briefly, I monitored the degradation of cefotaxime by TEM at 25°C via a change in absorbance at 262 nm using a Cary 100 UV–Vis spectrophotometer (Agilent Technologies) in 50 mM potassium phosphate, pH 7.0 with 10% glycerol. The final protein concentration was no greater than 200 nM. 5-200 µM cefotaxime was incubated at 25°C for 5 min before addition of the protein, and I measured each cefotaxime concentration in triplicate. When possible, the Michaelis-Menten equation (below)
was fit to the initial velocity \( (v_i) \) as a function of the substrate concentration \( [S] \) to determine individual catalytic rate \( (k_{cat}) \) and Michaelis constant \( (K_M) \) values. Here, \( [E] \) is the total enzyme concentration.

\[
v_i = \frac{[E] \cdot k_{cat} \cdot [S]}{K_M + [S]} \tag{3.4}
\]

When the \( K_M \) was too high to reach maximum velocity, a line (below) with a slope equal to \( [E] \cdot \frac{k_{cat}}{K_M} \) was fit to the data. Error for the fit parameters was determined using bootstrapping.

\[
v_i = [E] \cdot \frac{k_{cat}}{K_M} \cdot [S] \tag{3.5}
\]

### 3.5.3 Urea Denaturation Experiments
For both fluorescence and CD experiments, I prepared samples of 35 \( \mu \)g/mL protein in 50 mM potassium phosphate, pH 7.0 at various concentrations of urea and equilibrated the samples at room temperature overnight. After one-minute incubation in the instrument, I performed experiments as previously described [16, 18]. I recorded the signal at 25°C in a one cm path length quartz cuvette and measured the refractive indexes of each sample afterward in order to determine their precise urea concentration.

To monitor fluorescence \( (F) \) emission at 340 nm, I used a Jasco FP-8300 spectrofluorometer equipped with a Peltier-controlled cuvette holder. To monitor circular dichroism \( (CD) \) signal at 222 nm, I used an Applied Photophysics Chirascan equipped with a Quantum Northwest Inc. TC125 Peltier-controlled cuvette holder. I determined free energy values by fitting a two-state folding model (below) to the fluorescence data and a three-state folding model (below) to the CD
data using the linear extrapolation method [19]. Each variant was measured in triplicate experiments.

\[
F = \frac{F_U + F_N \times e^{-(\Delta G + m + \text{[urea]})/R + T}}{1 + e^{-(\Delta G + m + \text{[urea]})/R + T}}
\]  

(3.6)

Here, \(F_U\) and \(F_N\) are the fluorescence signals for the unfolded and native states, fit as lines. \(\Delta G\) is the extrapolated free energy difference between the unfolded and native states in the absence of urea, and \(m\) is the proportionality constant related to the steepness of the folding transition. \(R\) is the gas constant, and \(T\) is temperature.

\[
CD = \frac{\theta_U + \theta_I \times e^{-(\Delta G_U + m_U + \text{[urea]})/R + T} + \theta_N \times e^{-(\Delta G_U - m_U + \text{[urea]})/R + T} \times e^{-(\Delta G_I + m_I + \text{[urea]})/R + T}}{1 + e^{-(\Delta G_U + m_U + \text{[urea]})/R + T} + e^{-(\Delta G_I + m_I + \text{[urea]})/R + T}}
\]  

(3.7)

Here, \(\theta_U\), \(\theta_I\), and \(\theta_N\) are the CD signals for the unfolded, intermediate, and native states, fit as lines. \(\Delta G_{U \leftrightarrow I}\) and \(\Delta G_{I \leftrightarrow N}\) are the extrapolated free energy difference between the unfolded and intermediate states and the difference between the intermediate and native states, respectively. \(m_{U \leftrightarrow I}\) and \(m_{I \leftrightarrow N}\) are the proportionality constants related to the steepness of the two folding transitions. Again, \(R\) is the gas constant, and \(T\) is temperature.

For the pulse proteolysis experiments [20], I equilibrated samples of 0.5 mg/mL TEM in 20 mM Tris, pH 7.0 plus 10 mM CaCl\(_2\) and various concentrations of urea at room temperature overnight. Then, I added 0.05 mg/mL of thermolysin (Promega) to each tube and vortexed briefly. Thermolysin is active in urea concentrations up to 7 M and was stored in 2.5 M NaCl and 10 mM CaCl\(_2\) [20]. After 1 minute, I removed an aliquot of each reaction and added it to a tube with 50 mM EDTA to quench the proteolytic reaction. I then separated the samples using SDS-PAGE, stained the gel with SyproRed in 7.5% acetic acid, and imaged the gel using a Typhoon FLA 7000 biomolecular imager (GE). Finally, I quantified the gel band intensities using ImageJ.
(Fiji) and fit a two-state folding model (as shown above) to the proteolysis data, using the value for \( m \) determined from the fluorescence experiments.

### 3.5.4 Calculation of Fraction Folded

Fraction folded was calculated from \( \Delta G_{N\rightarrow I} \) using the equation below.

\[
\text{Fraction folded} = \frac{e^{-\frac{\Delta G_{N\rightarrow I}}{R+T}}}{1 + e^{-\frac{\Delta G_{N\rightarrow I}}{R+T}}}
\]  

\( \Delta G_{N\rightarrow I} \) is the free energy difference between the intermediate and native states determined from fluorescence experiments. \( R \) is the gas constant, and \( T \) is temperature.

### 3.5.5 Protein Abundance Using Western Blotting

First, I first generated TEM variants in the pBR322 plasmid using site-directed mutagenesis, verified the mutations via DNA sequencing, and transformed plasmids into DH5\( \alpha \) cells (Life Technologies). I also used these plasmids in the MIC experiments as they use tetracycline resistance for selection, the native TEM promoter for constitutive protein expression, and the native TEM signal sequence for periplasmic export. I grew 5 mL liquid cultures to saturation before harvesting cells and resuspending them in 20 mM Tris, pH 7.0 plus Roche protease inhibitors. Next, I lysed cells using freeze thaw cycles followed by sonication. I then quantified the total concentration of protein in each sample using a Bradford assay and made samples with equal amounts of total protein to separate via SDS-PAGE.

I transferred the protein to a polyvinylidene fluoride (PVDF) membrane (Bio-Rad) at 4°C in 25 mM Tris base, 192 mM glycine, 20% (v/v) ethanol. After transfer was complete, I equilibrated membranes first in methanol then in Tris buffered saline Tween-20 (TBST) supplemented with
5% (w/v) powdered dry milk (PDM) for 5 minutes. I added a 1:2,000 dilution of mouse β-lactamase primary monoclonal antibody (ThermoFisher) to the TBST/5% PDM solution and incubated 45 minutes while rocking. I washed the membrane with TBST three times to remove unbound antibody and then incubated for 45 minutes with a 1:10,000 dilution of rabbit anti-mouse IgG/IgM alkaline phosphatase (AP)-conjugated polyclonal secondary antibody (ThermoFisher) in TBST/5% PDM while rocking. Again, I washed the membrane three times with TBST to remove unbound antibody and then rinsed twice with deionized water. Finally, I developed the membrane by incubating it with alkaline phosphatase buffer (100 mM Tris HCl, pH 9.5, 100 mM NaCl, 5 mM MgCl2) containing nitro blue tetrazolium and 5-bromo-4-chloro-3-indoyl-phosphate (Promega), until bands turned dark purple. I stopped the reaction by rinsing with water and left the membranes to dry overnight before imaging.

### 3.5.6 96-well Plate Assay for Protein Abundance

I used the same plasmids, cells, and cell lysis protocol as the protein abundance measurements via Western blots described in the main text methods. Again, I measured the total cellular protein abundance using a Bradford assay. Instead of running a gel, however, I calculated the β-lactamase concentration in the cells by measuring the rate of hydrolysis of (300 µM) nitrocefin (Cayman Chemical) by the cell lysates. I included buffer only (50 mM potassium phosphate, pH 7.0) samples and a dilution series of purified protein in my 96-well plate. I next generated dilution series of cell lysates, of cells both with and without the plasmid containing β-lactamase. I then used a Synergy2 plate reader (BioTek) equipped with an injector in order to add an equal volume of nitrocefin to each well and immediately measure the change in absorbance at 482 nm for 30 seconds at 25°C.
To calculate the protein concentration, I first calculated the slope of the linear signal from each well of my plate. I then plotted the slope as a function of known protein concentration for my wells containing purified protein. This allowed us to obtain an in-plate determination of the relationship between absorbance signal and protein concentration, as the different variants have different catalytic rates for nitrocefin. I then divided the calculated slope from each of my cell lysate wells with the slope from my purified protein standard curve and converted this value to mg/mL. After accounting for the dilution factor for each well and normalizing by total cellular protein abundance, I report the average protein abundance and standard deviation for each experimental well within the dynamic range.

### 3.5.7 Minimal Inhibitory Concentration Measurements

I transformed plasmids into DH5α cells and grew 5 mL liquid cultures to saturation. I then diluted cells to an OD_{600} of 0.0003 in sterile Mueller Hinton II (MHII) media broth (Sigma). As previously [16], to determine the minimal inhibitory concentrations of cefotaxime, I used the broth microdilution method. I used water with no antibiotic in one row for a positive control for cell growth and used the antibiotic gentamicin in another row for a negative control for cell growth. I prepared 20 mM cefotaxime by dissolving the compound in water, prepared cefotaxime dilutions in MHII, and added these to the first well of each remaining row in a 96-well round-bottomed plate. I then made two-fold serial dilutions across each row. Finally, I added a volume of bacteria culture equal to the antibiotic volume in the plate to each well and incubated the plate at 37°C for 17 hours. I determined MIC as the lowest concentration of cefotaxime that inhibited cell growth as judged by visual inspection of turbidity.
3.6 References


Chapter 4: Conclusion and Future Directions

4.1 Excited States Containing a Cryptic Pocket Can Play a Functional Role

Cryptic pockets have been discovered in many different systems [1-3], but because they are difficult to observe, it has been unclear if cryptic pocket opening is generally conserved. While it has been proposed that cryptic pockets could provide a means to achieve homolog- or isoform-specific therapeutics [4], this has yet to be rigorously tested. In Chapter 2, we found that neither MTB nor GNCA opened a pocket in the region of the Ω-loop cryptic pocket in TEM, suggesting that the thiol labeling technique used is an unbiased method for searching for cryptic pockets. This result also shows that the Ω-loop cryptic pocket is not conserved across β-lactamase homologs despite the homologs sharing the same topology. However, the protein sequences were different, suggesting the importance of sequence in determining cryptic pocket opening.

We also found that homologs with Ω-loop pockets were able to better hydrolyze benzylpenicillin as compared to homologs without pockets. Furthermore, by focusing the remainder of our study on TEM β-lactamase, we found that variants with higher populations of open Ω-loop pockets were also able to better hydrolyze benzylpenicillin. These increases in catalytic efficiencies were mostly due to increases in catalytic rate, suggesting a role in Ω-loop dynamics in the hydrolysis mechanism. Combining these results with previous work that demonstrated a correlation between cefotaxime activity and the population of closed Ω-loop conformations [5], we hypothesized that an open Ω-loop pocket is beneficial for benzylpenicillin activity but detrimental to cefotaxime activity.
In order to gain structural insight, we examined the dynamics of TEM β-lactamase using NMR chemical exchange saturation transfer (CEST) experiments [6]. CEST allows us to observe excited states in proteins that have populations as small as 1% or less and does not require a cysteine mutation. Thus, these experiments can nicely corroborate findings from thiol labeling experiments. For TEM, we found that an excited state is present in the native ensemble and that the residues that contributed to the exchange between the ground and excited state were mostly in the region of the Ω-loop pocket. Furthermore, the population of the excited state determined from CEST experiments matched the population determined from thiol labeling, and when the Ω-loop pocket was removed from TEM, the dynamics monitored via CEST disappeared. We next ran NMR-guided molecular dynamics simulations that seeded unbiased simulations. This strategy allowed us to observe cryptic pocket opening and obtain detailed structures of the pocket open and closed states. Taken together, these results suggest that excited states containing cryptic pockets can be observed using thiol labeling, NMR CEST, and molecular dynamics, which expands our ability to investigate these excited states in any system.

Our model that an open Ω-loop pocket is beneficial for benzylpenicillin activity but detrimental to cefotaxime activity is in contrast with the hypothesis that cefotaxime activity requires an expansion of the active site due to the substrate’s increased size compared to benzylpenicillin [7]. If the latter hypothesis were true, one would expect that opening the Ω-loop pocket would increase cefotaxime activity and decrease benzylpenicillin activity while the opposite would be true for our proposed model. To explicitly test the effects of pocket dynamics on activity, we designed mutations which either enhanced or suppressed Ω-loop pocket opening. We found that
there was a clear positive correlation between Ω-loop pocket opening and benzylpenicillin activity, suggesting that an open Ω-loop pocket is beneficial for benzylpenicillin activity. Furthermore, we also found that a closed Ω-loop pocket is beneficial for cefotaxime activity, despite being a larger substrate.

In summary, our results suggest that excited states containing a cryptic pocket can play a role in protein function. This suggests that they may be favored by natural selection, making cryptic pockets more tractable drug targets. If a protein cannot easily evolve away cryptic pocket opening because it is important for function, it will be more difficult to evolve resistance to a drug which binds a cryptic pocket. The functional role of excited states containing a cryptic pocket also suggests that the use of pocket opening and excited states in general may be more prevalent in molecular mechanisms than we currently realize. Overall, this work demonstrates the utility of excited states containing cryptic pockets and provides evidence for their continued study.

4.2 Limitations of our Work Investigating Cryptic Pockets.

As with any scientific study, there are limitations to our analyses and interpretations. For example, our analysis of cryptic pocket conservation was limited to one representative sequence from each β-lactamase homolog family. So, we cannot say if other MTB sequences in fact do contain the Ω-loop pocket or if other TEM and CTX-M sequences do not contain the Ω-loop pocket. Furthermore, our analysis was limited to one cryptic pocket in class-A β-lactamases. Future studies may investigate the conservation of cryptic pockets in other systems and using a larger number of sequences. Additionally, our labeling assays provide a quantitative description
of pocket opening but do not provide a full characterization of the protein energy landscape. Thus, there may be important protein motions that contribute to activity that are not visible by monitoring cysteine labeling. These dynamics could be larger scale motions such as surface loop movements or smaller scale motions such as different sidechain rotamers. As a consequence, the population of the open cryptic pocket conformations is correlated with benzylpenicillin activity and is a predictive signature, but we cannot say that there is a causal link between the two.

4.3 Future Studies to Elucidate the Functional Role, Sequence Determinants, and Cooperativity of Cryptic Pocket Opening

4.3.1 Elucidate the Sequence Determinants of Cryptic Pocket Opening

We found that four β-lactamase homologs with the same topology but different sequences had different open Ω-loop cryptic pocket populations. Despite sharing the same overall fold and containing the Ω-loop structural motif, two of the class-A β-lactamase homologs, MTB and GNCA, did not open a pocket in the region of the Ω-loop pocket in TEM and CTX-M-9. Furthermore, a singular R241P mutation was able to remove the Ω-loop pocket in TEM completely. These results suggest sequence is important for determining cryptic pocket opening, but it is still unclear what the sequence determinants for pocket opening are.

Figure 4.1. TEM and GNCA β-lactamases share 67% sequence identity in the region of the Ω-loop pocket. The regions of the Ω-loop and 238-loop are underlined. The positions where the two sequences differ are highlighted in yellow.

GNCA does not have an Ω-loop cryptic pocket but is the ancestor sequence for TEM β-lactamase [8], which has the highest population of open Ω-loop pocket for the homologs tested.
These two homologs also have similar sequences around the Ω-loop pocket region, sharing 67% sequence identity and 78% sequence similarity (Figure 4.1). To better understand the sequence determinants of cryptic pocket opening, I propose determining the minimal set of TEM-like mutations needed to open the Ω-loop pocket in GNCA.

There are six TEM-like mutations that could be made in GNCA in the Ω-loop region: A173I, G175N, P177E, G240E, H241R, and T243S (Figure 4.1). I propose incorporating each mutation into GNCA individually and testing the effects of the mutation on thiol labeling. The two mutations with the most potential are P177E and G240E. P177E would remove a proline residue in the Ω-loop of GNCA, which would likely increase its conformational freedom. We found that adding in a proline residue into the Ω-loop pocket of TEM (R241P) removed the pocket, so I reason that the opposite is likely true for GNCA. However, there is a proline residue at position 177 in CTX-M-9 that does open an Ω-loop pocket, suggesting that broader sequence context may also be important.

G240E, on the other hand, would add a negatively charged residue in the Ω-loop region of GNCA right across from the glutamic acid at position 171. I reasoned that this may destabilize the closed pocket conformation and promote pocket opening. However, I did test this mutation and found that the Ω-loop pocket does not open in GNCA G240E (Figure A.4).

If a single mutation is unable to open the Ω-loop pocket in GNCA, mutations could be incorporated in combination. However, with 6 mutations, there are $2^6 = 64$ possible combinations. Using the data from Stiffler et al. [9], we can see which of the GNCA-like
mutations in TEM increased fitness in the presence of cefotaxime. These mutations likely close the Ω-loop pocket in TEM, so the opposite mutations might open the Ω-loop pocket in GNCA. The two mutations that meet this criterion are A173I and G240E, suggesting the prioritization of the single mutant A173I and the double mutant A173I/G240E. Alternatively, molecular dynamics simulations, and particularly those run with the enhanced sampling algorithm FAST pockets [10], could be used to select variants to experimentally test.

Elucidating the sequence determinants of cryptic pocket opening has important implications in protein design as it would increase our understanding of how mutations change not only the ground structure of a protein but its relative populations of excited state structures. As we showed in Chapter 2, protein excited states can be functionally relevant, and future design efforts will likely design towards an ideal protein energy landscape rather than one ideal protein ground state structure [11]. Understanding sequence determinants will also be useful in the prediction of which homologs or isoforms contain cryptic pockets and would allow for the design of more specific therapeutics that target cryptic pockets [4, 12].

### 4.3.2 Determine if the Open Ω-loop Pocket Population is Correlated to an Individual Rate Constant in the β-lactamase Catalytic Mechanism

We found that β-lactamase homologs which open an Ω-loop pocket have higher overall catalytic rates (k\text{cat}) against benzylpenicillin than homologs which do not open an Ω-loop pocket. We also found that for TEM variants, the open Ω-loop pocket population is correlated with the catalytic rate for hydrolysis of benzylpenicillin. However, TEM-1 β-lactamase hydrolyzes benzylpenicillin with an efficiency near the diffusion limit, meaning that the individual rate
constants for the reaction are fast and comparable to one another [13]. This leads to an overall catalytic rate that is described by multiple rate constants.

Hydrolysis of β-lactams by β-lactamase follows the reaction scheme shown below.

\[
E + S \overset{k_1}{\underset{k_{-1}}{\leftrightarrow}} ES \rightarrow ES^* \rightarrow E + P
\]  

(4.1)

\(E\) is the enzyme (β-lactamase), \(S\) is the substrate (β-lactam), \(ES\) is the enzyme-substrate complex, \(ES^*\) is the acyl-intermediate, and \(P\) is the product (hydrolyzed β-lactam). \(k_1\) and \(k_{-1}\) are the binding and unbinding rates, respectively. \(k_2\) is the acylation rate and \(k_3\) is the deacylation rate. From this scheme, the expression for the catalytic rate can be derived as the following equation [14].

\[
k_{\text{cat}} = \frac{k_2 \times k_3}{k_2 + k_3}
\]  

(4.2)

The Ω-loop (and particularly E166) is necessary for the deacylation \((k_3)\) of β-lactams, and deacylation is the rate-limiting step for benzylpenicillin hydrolysis [15]. However, the acylation rate \((k_2)\) is the rate-limiting step for cefotaxime hydrolysis [16] and has been proposed to be the distinguishing feature between the hydrolysis of penicillins and cephalosporins [15]. Thus, I hypothesize that the open Ω-loop pocket population is correlated with the acylation rate.

Single turn-over and steady-state kinetics can be used to monitor the hydrolysis of slow substrates by β-lactamases in order to determine the individual acylation and deacylation rates [14, 16]. However, determining the rates for β-lactamases hydrolyzing good substrates like benzylpenicillin is more challenging. Neither single-turnover kinetics in a quenched flow (there
is too much overlap between the benzylpenicillin and enzyme absorption) nor the stopped flow (the reaction is too fast) are possible [15]. The acylation and deacylation rates instead can be determined using a quenched flow equipped with three syringes to quantify the fraction of acyl-enzyme ($ES^*$) [15]. To date, all of the rate constants have been determined for β-lactamase I, PC1, and RTEM hydrolyzing benzylpenicillin as well as for β-lactamase I and PC1 hydrolyzing other substrates [17].

I propose performing DTNB-labeling on TEM E104K and TEM E104K/G238S, as these variants were shown computationally to have different populations of open Ω-loop pocket conformations [5]. These variants close the Ω-loop pocket and have catalytic rates against benzylpenicillin that are correlated with their open Ω-loop pocket conformations. Combined with the data for wild-type TEM, DTNB-labeling experiments will corroborate the computational data as well as provide a quantitative measure of the pocket open population. Alternatively, the Ω-loop pocket populations could be used for wild-type TEM, TEM E240D, and TEM E240D/R241P. TEM R241P will likely not be useful for this study as it removes the Ω-loop pocket entirely and has a higher catalytic rate against benzylpenicillin as compared to TEM E240D/R241P.

Next, I propose determining the acylation and deacylation rates of the variants of interest for benzylpenicillin using the quenched flow measurement of the acyl-enzyme and for cefotaxime using the stopped flow, single turn-over kinetics methodology. I expect a negative correlation between the acylation rate (against both benzylpenicillin and cefotaxime) and the population of open Ω-loop pocket, meaning that closing the Ω-loop pocket increases the acylation rate. This increased acylation rate would be especially beneficial for cefotaxime hydrolysis, as the wild-
type enzyme has a slow acylation rate. For benzylpenicillin hydrolysis, an increased acylation rate likely comes at the cost of a decreased ability to bind and release product and thus a corresponding decrease in overall catalytic efficiency. If needed, the binding and unbinding rates can be determined by measuring the catalytic efficiencies as a function of viscosity [17].

We showed in Chapter 2 that an open cryptic pocket population is correlated with protein function, as measured by catalytic efficiency. Specifically, the changes seen in catalytic efficiency are due to changes in catalytic rate. Using NMR and molecular dynamics simulation data, we also proposed a model for how the open and closed pocket conformations might affect the hydrolysis reaction, which the experiments proposed above would more explicitly test. These experiments would also add to our understanding of how cryptic pocket opening in enzymes is connected to enzyme activity. The general principles learned in these studies, particularly when combined with the work proposed in the previous section, would aid in protein design.

4.3.3 Determine the Cooperativity of the Ω-loop Pocket and the Helix 11/12 Pocket in TEM β-lactamase
We found that the dynamics of the Ω-loop pocket in TEM were able to affect the active site residues and ultimately activity against different classes of substrates. Besides the Ω-loop pocket, there is another known cryptic pocket in TEM β-lactamase between helices 11 and 12, and we call this the helix 11/12 pocket (Figure 4.2). As this pocket was found serendipitously during a drug screening campaign, it is known that this cryptic pocket has the ability to allosterically affect the active site [18]. In fact, several allosteric modulators have been discovered by targeting this pocket [19]. Both cryptic pockets individually have allostery with the active site, but do they allosterically communicate with one another?
Figure 4.2. TEM β-lactamase has an additional cryptic pocket termed the helix 11/12 pocket. The structure shown is the crystal structure (PDB: 1pzo) where an inhibitor was found to bind into a pocket absent in the ligand-free crystal structure. The inhibitor structure has been removed from the pocket here in order to highlight the pocket opening. The regions of the Ω-loop pocket and helix 11/12 pocket are highlighted with dashed circles.

The residues that were found via NMR CEST experiments to participate in the exchange between the ground state and open cryptic pocket excited state were those in the region of the Ω-loop pocket as well as some residues within the β-sheet between the Ω-loop and helix 11/12 pockets. We also saw a similar behavior from the CARDS analysis on TEM where the Ω-loop community continues into the β-sheet. These results suggest that the two pockets communicate with one another, perhaps through the β-sheet. Communication of distal sites via a β-sheet is an emerging strategy of allosteric communication [20]. As each open cryptic pocket structure has been observed computationally [21-23], but there is currently no structure of both pockets open at the same time, I hypothesize that the two pockets are negatively coupled in that opening one pocket closes the other.

To test this hypothesis, I propose using the L286C mutation in TEM to monitor opening of the helix 11/12 pocket and then comparing the labeling of the helix 11/12 pocket in both wild-type
and R241P. I expect that the R241P mutation, which closes the Ω-loop pocket, will increase the population of open helix 11/12 pocket as compared to wild type. Previously, the pocket opening rate of the helix 11/12 pocket was determined as the labeling was in the EX1 regime [22]. However, lower DTNB concentrations could push the labeling to the EXX regime, making the determination of the pocket population possible.

Less is known about the helix 11/12 pocket in terms of its conservation across β-lactamases, connection to activity, and mutations which enhance/suppress pocket opening. Preliminary data suggest that unlike the Ω-loop pocket, the helix 11/12 pocket is not conserved in CTX-M-9 (Figure A.3). It would be interesting to know if GNCA contains a helix 11/12 pocket, as this would suggest that the Ω-loop pocket was gained in an ancestor of TEM and CTX-M-9 but then the helix 11/12 pocket was lost in an ancestor of CTX-M-9. This could be tested by incorporating a cysteine residue at position 286 in GNCA and monitoring labeling. Alternatively, the helix 11/12 pocket might only exist in TEM β-lactamase, making the reason for its allosteric connection to activity unclear. Additionally, the helix 11/12 pocket could be used to gain understanding of the sequence determinants of cryptic pocket opening, using a similar strategy to the one outlined in a previous section. Then, mutations identified to close the helix 11/12 pocket in TEM could be used to test the effects of closing this pocket on the open population of the Ω-loop pocket. Again, I would expect that closing the helix 11/12 pocket would open the Ω-loop pocket. Though, to my knowledge, the reversibility of allosteric communication has yet to be tested.
Understanding the connectivity of multiple cryptic pockets adds to our understanding of allostery in proteins, and the existence of multiple connected pockets increases the therapeutic potential of a protein. Finding specific and tight binders even to known and well-defined pockets can be difficult but having multiple pockets in a protein that are able to affect the active site would increase the potential of success. Particularly, if drug binding to a pocket undesirably perturbs an important residue, modulating the dynamics of that pocket via drug binding to an allosterically connected pocket presents a new therapeutic opportunity.

4.4 Considering Excited States Along the Protein Folding Pathway Increases our Ability to Predict Fitness from Protein Properties

While the catalytic efficiency of an enzyme typically well-describes its function, in Chapter 3, I found that the catalytic efficiency of TEM β-lactamase variants alone only explained about 70% of the variance in bacterial fitness, as measured by minimal inhibitory concentration (MIC). Strikingly, two variants with similar activities fell far off of the 1:1 correlation line in opposite directions. This suggests that one or more protein properties also significantly contribute to fitness.

It has previously been suggested that because TEM variants are very stable, particularly with fraction folded values above 99%, stability values add no predictive power [24]. However, the authors of this study assumed a two-state model to describe TEM folding/unfolding. A two-state folding model determines the free energy difference between the ground state and the unfolded state and is often used because of the cooperative nature of protein folding. If a folding
intermediate is not significantly populated or does not produce a signal different from the ground or unfolded states, than it cannot be seen in bulk, *in vitro* measurements of stability. In order to evaluate the validity of the two-state assumption, protein stability is measured using multiple, independent signals. If the transitions for each normalized curve do not overlay, this suggests that there are more than two states. That being said, I found that three states (the ground state, the intermediate state, and the unfolded state) can be observed from a single stability experiment monitored using circular dichroism (CD) at 222 nm.

While the intermediate state is a protein excited state with a lower energy (higher probability) than the unfolded state, I found that the free energy differences between the ground and intermediate states still correspond to fraction folded values above 99%. However, I hypothesized that destabilization that leads to partial unfolding would make the protein susceptible to aggregation and/or degradation in the cellular environment. By measuring stability using a pulse proteolysis assay, I found that the TEM folding intermediate is in fact susceptible to proteolytic digestion. By measuring protein cellular abundance, I also found that destabilized variants (*i.e.*, those with larger concentrations of the folding intermediate) have drastically reduced concentrations in the cell. These results taken together provide evidence for the importance of stability in determining fitness and suggest that *in vitro* measurements of stability may add predictive power, though not expressed as fraction folded.

Next, I derived an expression to describe how stability would alter the concentration of the ground state conformation by first considering a kinetic scheme in which the protein could be synthesized to the intermediate state (I assumed the concentration of the unfolded state to be
negligible) and then fold to the ground state. I also considered that the protein could unfold back
to the intermediate state and be proteolyzed from there. Using catalytic efficiency values,
stability values, and the resulting model, I was then able to predict 99% of the variance in MIC.
In order to achieve this correlation, I applied a stability threshold based on the observation that
stability past a certain point does not increase cellular abundance or fitness. My results are in
excellent agreement with previous work [25] that suggested that stability accounts for about 29%
of the variance in MIC and that TEM is only about -1.7 kcal/mol stable in vivo.

In summary, I found that both protein activity and stability contribute to bacterial fitness.
Sufficient catalytic efficiency is needed to order to degrade β-lactam antibiotics before they
reach their targets, but increases in catalytic efficiency often come at the cost of stability.
Because the cellular environment contains protein quality control machinery, like proteases,
decreases in stability can lead to decreases in protein cellular abundance. This is particularly true
because a protein samples excited state conformations, and a protein does not need to unfold
entirely in order to aggregate or be degraded. So, while it can still be beneficial to have a
decreased number of highly active enzymes (as I saw with TEM E104K/G238S), stability is an
important consideration for the quantitative prediction of fitness.

4.5 Limitations of my Model for Predicting Fitness.
As with any model, my model is not appropriate to use in all situations. To start, I used a
measure of fitness, minimal inhibitory concentration of an antibiotic, that only gives a measure
of the protein’s fitness in the presence of antibiotic. In the absence of antibiotic, the activity of
TEM becomes nonessential, though the stability of the protein remains important. The bacteria
may even have a lower tolerance for destabilized variants that provide no useful function. In deriving my model, I also assume that the synthesis and proteolysis rates are independent of the TEM variant being produced. As the variants differ by only a couple of nucleotides, the synthesis rate is unlikely to be drastically affected, though this was not explicitly tested. The proteolysis rates could more likely change due to a single amino acid change if the resulting mutation introduces a cut site for the native protease. Furthermore, the synthesis and proteolysis rates become increasingly important with an increase in sequence diversity. Finally, it is possible that addition of stability in my model fits to the data very well trivially due to the addition of parameters. The validation of the model on a new set of variants not used to fit the model provides evidence that this is not the case, but it is difficult to determine if a model is overfit without additional data.

4.6 Future Studies to Validate my Model for Predicting Fitness

4.5.1 Test the Generalizability of the Model by Using Different Classes of Antibiotics
I found that my model explained 99% of the variance in the MIC of the third-generation cephalosporin, cefotaxime. I expect that my model is generalizable to other antibiotics as the protein factors which contribute to fitness are unlikely to change with the class of antibiotic, but this has yet to be tested. Activity is dependent on the antibiotic tested, but I do not expect stability to be. For substrates which are slowly hydrolyzed, there may be an appreciable amount of time in which the bound substrate stabilizes the enzyme. However, I did not find this to be the case for the slowly hydrolyzed substrate cefotaxime.
Thus, I hypothesize that measuring the TEM variants’ activity against benzylpenicillin (a penicillin) and imipenem (a carbapenem), combined with known stability data, will be predictive of MIC of the respective antibiotics (Figure 4.3). The $Z'$ parameter will need to change as the permeability of the antibiotic through the outer membrane is included in this parameter and is an antibiotic-specific value. However, because this is a known quantity, the originally fit $Z'$ parameter can simply be scaled accordingly.

These experiments would be beneficial as the more generalizable the model is, the more useful it can be. Application to different classes of antibiotics also suggests a good understanding of the protein properties which contribute to fitness and that only two parameters are needed to describe factors independent of the protein variant being produced.

4.5.2 Determine Which Protease(s) Degrade TEM $\beta$-lactamase in vivo
I found that destabilized TEM variants had drastically reduced cellular protein abundance due to proteolytic digestion from a TEM excited state. I tested the proteolytic susceptibility of the
folding intermediate population using the protease thermolysin and fit a parameter which contained an intrinsic proteolysis rate. However, thermolysin is an extracellular metalloendopeptidase secreted by the Gram-positive, thermophilic bacterium Bacillus thermoproteolyticus [26] and is thus unlikely to be the protease that degrades TEM intracellularly in Gram-negative Escherichia coli. This leads to the question: which protease is responsible for degrading TEM β-lactamase in E.coli?

There are 10 known periplasmic proteases: degP, degQ, ompT, ptrA, tsp, ycaL, ydgD, loiP, bepA, and yhiJ [27]. Of these, five proteases (ompT, ptrA, tsp, loiP, and bepA) are unlikely to degrade TEM based on their known substrates, and three proteases are uncharacterized (ycaL, ydgD, and yhiJ). The most likely proteases that degrade TEM are DegP, which degrades misfolded proteins [28] and DegQ which degrades partially unfolded substrates [29]. Thus, I hypothesize that TEM is degraded in the periplasm by DegP and/or DegQ.

To determine if these proteases do in fact degrade TEM in vivo, I propose performing the same cellular protein abundance assays from Chapter 3 but with cells lacking each protease. I expect that when the native protease is absent, there will be increased cellular abundance of the destabilized TEM variants, particularly in the whole cell lysate and insoluble fractions. If neither of these proteases alter TEM abundance, either alone or in combination, the three uncharacterized periplasmic proteases could be tested next.

Determining the native protease which degrades TEM would allow for the measurement of the intrinsic proteolytic rate of that enzyme, a value currently fit within the $Z'$ parameter in my
model. The more information we have about the biological context of TEM and specifically the more quantitative information we have about the factors independent of the TEM variant being produced, the better we can predict bacterial fitness from the biophysical properties of TEM alone. Additionally, knowing the native protease would allow us to determine if the folding intermediate seen during in vitro measurements of stability is the state which gets degraded by this protease (as it is for thermolysin). This would provide further insight into the protein energy landscape of TEM β-lactamase and provide further evidence for the functional importance of excited state conformations.

4.7 Closing Thoughts

It is well-established that proteins are dynamic molecules, but studying the multiple conformations a protein adopts and relating their structures and probabilities to the protein’s function remains a challenge. However, proteins are nanomolecular machines and, like any machine, if they do not move, they cannot function. Thus, the questions surrounding protein dynamics are difficult but extremely important. In recent years there have been significant advances towards understanding how proteins move and an increased appreciation for those movements. This thesis specifically emphasizes the utility of considering the excited states of proteins in protein design, drug discovery, and elucidating the molecular mechanisms of disease.

4.8 References


Appendix

A.1 Figures

Figure A.1. One cysteine labels in both wild-type (WT) MTB and MTB A285C. The normalized DTNB labeling traces for WT MTB (orange circles) and MTB A285C (light orange pentagons) plateau at one cysteine labeling. We know from Chapter 2 that the cysteine which labels in WT MTB is C287. We introduced the A285C mutation in order to test for the presence of the helix 11/12 pocket in MTB. While only one cysteine labels in both variants, it is unclear if the difference in labeling rate is due to the A285C mutation increasing the solvent exposure of C287 or if the introduced cysteine itself labels and causes C287 to no longer be accessible.
Figure A.2. DTNB labeling WT MTB inhibits activity against nitrocefin. Michaelis-Menten kinetics were measured for unlabeled wild-type MTB (orange solid line) and TNB-labeled wild-type MTB (dark orange dashed line). Error bars represent the standard deviation of three independent runs. We know from Chapter 2 that the cysteine which labels in wild-type MTB is C287. This data suggests allosteric communication between the region of position 287 and the active site. Interestingly, the differences in catalytic efficiency are due to changes in both the $k_{\text{cat}}$ and $K_M$. 
Figure A.3. Both cysteines label in CTX-M A283C. (Left) The normalized DTNB labeling trace for wild-type (WT) CTX-M (cyan circles), as expected, plateaus at one cysteine labeling. However, the labeling trace for CTX-M A283C (light cyan diamonds) plateaus at a value greater than one cysteine labeling and is well-fit by a double exponential. I introduced the A283C mutation in order to test for the presence of the helix 11/12 pocket in CTX-M. (Right) I continued the analysis by measuring the observed labeling rate as a function of DTNB concentration and find that the slower cysteine (pentagons) labels on the order of magnitude of what is expected for protein unfolding. Error bars represent the standard deviation of three independent runs. I conclude that only one cysteine labeling corresponds to pocket opening in CTX-M A283C. However, it is still unclear if the faster labeling cysteine (diamonds) is at position 69 or 283. Either way, this data provides evidence for the allosteric communication between these two regions in CTX-M.

Figure A.4. The G240E mutation does not open the Ω-loop pocket in GNCA. The normalized labeling trace for GNCA G240E (a TEM-like mutation) shows no cysteine labeling.
Figure A.5. DTNB labeling of wild-type CTX-M results in a faster proteolysis rate while the presence of the compound NCI147639 results in a slower proteolysis rate. I measured the proteolysis rate (by thermolysin) of wild-type CTX-M unlabeled (cyan solid line), TNB-labeled (dark cyan dashed line), and in the presence of the compound NCI147639 (dark gray dash dotted line). TNB-labeling CTX-M is known to inhibit activity, presumably due to sterically blocking the Ω-loop [1]. Labeling also increases the proteolysis rate by over an order of magnitude. Compound binding, on the other hand, is known to stabilize proteins [2]. Addition of the NCI147639 compound slowed the proteolysis rate of wild-type CTX-M two-fold.

Figure A.6. TEM β-lactamase folding is not reversible by thermal denaturation. Fluorescence signal as a function of temperature was measured by first increasing temperature (lime green) to measure unfolding and then decreasing temperature (dark green) to measure refolding. The loss of signal in the refolding trace indicates that wild-type TEM does not reversibly fold after temperature denaturation.
A.2 References
