Activation and Inhibition of Biological Function through Design of Novel Protein-Ligand Interactions

By

Jimmy Budiardjo

Submitted to the graduate degree program in The Center for Computational Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson: Joanna Slusky, Ph.D.

Christian Ray, Ph.D.

Yinglong Miao, Ph.D.

Mark Richter, Ph.D.

John Karanicolas, Ph.D.

Eric Deeds, Ph.D.

Date Defended: 2 May 2019

The dissertation committee for Jimmy Budiardjo certifies that this is the approved version of the following dissertation:

Activation and Inhibition of Biological Function through Design of Novel Protein-Ligand Interactions

Chairperson: Joanna Slusky, Ph.D.

Date Approved: 2 May 2019

ABSTRACT

Virtually every process within a cell involves a protein. They serve as cellular workhorses carrying out functions such as catalysis of essential metabolites, to regulating which genes get turned on or off, to forming the structural scaffolding to retain rigidity of a cell. Proteins form the link between the genetic information encoded in DNA to the observable phenotype of an organism. The way proteins communicate is by direct physical contact with another molecule that alters its shape and dynamics to carry out a particular function. For example, G protein-coupled receptors are membrane imbedded proteins that bind to a small molecule or peptide in the extracellular environment and translate the binding event into an internal signal to regulate processes such as heart rate and even mood. The ability to selectively modulate such fundamental systems offers huge potential with broad applications from the ability to interrogate unknown cellular mechanisms to developing therapeutics when these interactions become aberrant.

The scope of this dissertation encompasses determining what properties dictate protein-ligand interactions and the application of these principles to the design of new ones. In particular, chapter 1 covers the design of a molecular switch that is turned *on* by small molecules. I follow this up in chapter 2 by investigating how to turn *off* protein function with small molecules in aberrant disease states. In chapter 3 we expand from the world of small molecule ligands to design a protein to turn *off* function of a protein involved in bacterial pathogenesis.

In loving memory of my mother, Tuty Budiardjo (1954-2016)

ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude to the two advisors I had during my time as a graduate student: Dr. John Karanicolas and Dr. Joanna Slusky. It has been a privilege to have had the opportunity to work under your guidance. Truly two of the young talents in your respective fields! John, your positivity and "You can do this!" attitude has helped me build confidence not only in the lab but also personally. Joanna, I am forever indebted to you for welcoming me into your lab during a difficult time. It has been exciting to witness and be a part of the growth and success of the lab. Thank you both for showing me patience during times when I have felt overwhelmed and for providing me with the tools necessary to develop into a productive scientist.

Additionally, I would like to thank my committee members: Dr. Eric Deeds, Dr. Wonpil Im, Dr. Mark Richter, Dr. Christian Ray and Dr. Yinglong Miao. Thank you for your time and feedback on my projects during seminars and for your open-door policy if there was anything I needed. Dr. Teruna Siahaan, thank you for your sincere concern for my professional development and for setting up my internship. Living in San Francisco for a summer and having the opportunity to work at such a prestigious place was an invaluable and incredibly enriching experience.

I want to thank all of my colleagues past and present for the commiseration over failed experiments and celebration of the successful ones! I am especially thankful to Dr. Yan Xia for helping me during the early years of graduate school. To Dr. Andrew Beaven for getting me introduced to the world of molecular dynamics simulations. To Jakki Deay for helping me finish

strong with project “get Jimmy graduated”. I appreciate all of your camaraderie over the years and I have truly made lifelong friendships.

To my Kansas City crew: Tiffany, Lance, Alyssa, Joe, Dani, Megan, Chet and Katelyn. Thanks for always being ready to grab a beer. To my out of state bothers: Nick, Sam, Alex and Ryan. We’ve come a long way since day one! To my car-wrenching crew: Andy, Alfonso, and Vince. There hasn’t been a better way to destress than spending a Sunday wrenching on cars.

I would not have been able to succeed without the support of my loving family. Special shout out to my puppies Norman and Maggie. Tamy, Cintya and Justin thank you for helping me get through this process in all of the *countless* ways that you did. Pam and Phil Schweisberger, thank you for embracing the Budiardjos and always keeping me well fed! At last, I want to thank my parents Agus and Tuty Budiardjo. I hope you know that your sacrifices to come to this country and make a better life for your children has paid off! Mom, I only wish you were here to see this and celebrate with us. I love you all!

TABLE OF CONTENTS

INTRODUCTION

Activation of protein function with small molecules - Building molecular switches

In the first chapter of this dissertation, I explore the rational design of a protein switch by engineering small molecule dependent control into an enzyme. A longstanding goal in chemical biology has been to build switches for diagnostics, biosensors, and synthetic biology. Small molecule dependent activity has been used to control pathways such as phosphorylation and protein degradation [1, 2]. Typically, this has been achieved using an engineering-centric bottom up approach where known functional domains are fused to a particular protein of interest to incorporate a regulatory domain not present in the wildtype protein [3]. For example, maltose-binding protein was fused to β-lactamase such that binding of maltose results in a positive modulator of β-lactamase activity [4]. Predicting how a functional domain transduces an allosteric signal into its fusion partner poses a difficult problem. Finding successful designs requires a lot of screening and trial and error especially related to insertion sites and linker length and composition. Instead of fusion of an "input domain" (such as maltose-binding protein) to an "output domain" (such as β-lactamase) we explored an alternative approach to engineer a new small molecule binding site directly in the output domain. This removes the necessity to have to distinct entities communicate and provides a more rational way to engineer control into the protein.

In order to achieve this, we devised a method called "chemical rescue of structure" [5-8]. First, we identify a buried "buttressing" side chain that is responsible for maintaining structural

integrity of the functional site of the protein (active site or binding site). We then remove the side chain (e.g. W→G) causing a cavity forming mutation that renders the protein inactive by disrupting the structural integrity of the active site. Addition of an exogenous ligand that can replace the structural moiety removed by mutation and restore structure and thus function.

As a proof of concept, we applied this to the sugar metabolizing enzyme β-glycosidase from a species of thermophilic archaeon. In this example a buried tryptophan residue next to the active site was mutated into a glycine. The resulting cavity caused an active site tryptophan reside to rotate into the cavity to fill the void. This alternate conformation breaks an essential hydrogen bond formed between the pyrrole ring nitrogen and hydroxyl group of the substrate sugar. Addition of exogenous indole binds the cavity and causing the active site tryptophan to rotate back into the catalytically competent conformation. The mechanism of disruption and subsequent reactivation of activity is mediated by a discrete conformational change in a single tryptophan residue.

In our proof of concept model, the cavity forming mutation was made by deleting a single tryptophan to glycine. This mutation caused a 730-fold reduction in k_{cat}/K_m which was fully rescued by complementing with indole. Indole in this case would replace all of the deleted atoms except for the β-carbon of tryptophan. Would replacing the β-carbon by adding a methyl group to the 3 position of indole improve rescue? In chapter 1, I explore the rules that dictate what makes the best protein-ligand pair such that these basic principles can be extended and applied to other systems [7]. What is the best way to design the cavity and complement it? How malleable is the *de novo* binding site? Will the effector-binding site be able to accommodate other ligands? To answer these questions, I applied a structure-activity relationship approach by exploring the

extent of rescue for a range of indole analogues. Additionally, I explore the possibility of using a double mutation to create a larger cavity.

Inhibition of protein function with small molecules - Exploring small molecule druggable states in conformational ensembles of protein-protein interfaces

In chapter 2, I use computational simulations to study the conformational landscapes of protein-protein interfaces (PPIs) to explore their druggability with small molecule inhibitors. Historically, targets of small molecules have been proteins whose cognate partners are themselves small molecules. These include enzymes, GPCRs, ion channels and transporters [9]. However, there is an ever-increasing interest in multiple classes of proteins that do not naturally bind small molecules for therapeutic intervention. These include proteins that bind to nucleic acids or to other proteins. In particular, systems involving protein-protein interactions have been much more difficult to target than traditional drug targets. Because they did not evolve to bind small molecules, they lack the deep pockets necessary interact with small molecules with high potency. The interaction surfaces between two proteins instead tend to be large and flat. This makes identification for potential sites of inhibition much more difficult.

In cases where inhibitors have been identified, a pocket forms on the surface that was not present in the unbound structure. Additionally, small molecule inhibitors do not need to span the entirety of the interface. It was found that PPIs interact through binding “hot spots” which are residues that disproportionately contribute to the majority of the binding free energy [10, 11]. Together, this shows that PPIs are indeed druggable. However, since the existence of these pockets is transient in nature, identifying the druggable states poses a difficult challenge.

Typical structure based experimental techniques like x-ray crystallography and NMR can reveal high-resolution snapshots of low energy states within the conformational ensemble. However, the dynamic nature of cryptic pockets means that transient states may not be captured using these techniques. Additionally, the pocket formed when bound to a small molecule is often different from the conformational changes observed when the protein is bound to its cognate protein partner. In recent years, the search for druggable cryptic pockets has been the focus of many computational efforts in an attempt to model them [12-14].

Molecular recognition of protein-ligand complexes falls under two models. In the induced fit model, the small molecule ligand causes the conformational change necessary to bind. In the conformational selection model, the conformational change happens before the ligand is present as part of the normal conformational fluctuations. The small molecule then binds to the preformed pocket and pushes the equilibrium towards that state. Computational methods like molecular dynamics simulations provide a robust tool for sampling conformational landscapes that may reveal the underlying mechanisms at play.

Once such tool has been developed in the lab using a biased sampling protocol within the Rosetta software suite [14]. Here we bias the conformational landscape to generate pocket containing conformations. If the conformational selection model holds true, our hypothesis is that pockets will form that are similar in shape to the pockets of known inhibitor bound structures. In chapter 2, I use this sampling method to explore the pocket dynamics within a set of four validated PPI targets. I then apply a high-resolution pocket shape comparison protocol that maps pocket topography of each snapshot and quantitatively assesses their shape similarity relative to known inhibitor bound pockets and between the ensembles themselves. The ultimate

goal is to then identify particular druggable transient pockets which can then be subjected to conventional computation aided drug design pipelines.

Inhibition of protein function with other proteins - Designing a molecular plug of outer membrane protein efflux pump TolC

The rise in multidrug-resistant bacteria has led to an increased unmet need for new therapeutics and strategies to combat deadly infections. The Centers for Disease Control estimates that in the United States alone, 2 million people will get antibiotic resistant infections resulting in the death of 23,000. Part of the issue is that there has been a dearth of new antibiotic drugs that have made it to market in the last several decades. The difficulty in bringing new drugs to market is a multifaceted complex problem that includes scientific, economic, and regulatory hurdles [15]. Multidrug-resistant clinical isolates deemed the ESKAPE pathogens (Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterobacter spp.) have been responsible for an increase in hospital acquired infections [16]. The ESKAPE acronym aptly designated for their ability to escape available antibiotic treatments. Of the seven pathogens, four are Gram-negative species. Gram-negative bacteria are particularly difficult to target due to their double-membrane cellular architectures that make a formidable barrier to penetrate. Furthermore, a new class of antibiotics targeting Gram-negative bacteria has not been introduced in the last half century [17]. One promising avenue however is the repurposing of currently available antibiotics to circumvent difficulties in generating new drug leads.

The increased ability to tolerate antibiotics is largely in part due to overexpression of broad-spectrum efflux pumps like the resistance-nodulation-division (RND) superfamily that extrude drugs and reduce their effective intracellular concentrations [18, 19]. Of the RND

superfamily, the most notable and heavily studied system is the AcrAB-TolC efflux pump from Escherichia coli. The AcrAB-TolC complex spans the entirety of the cell envelope from inner to outer membrane. AcrB forms the integral inner membrane transporter powered by proton motive force and contains the primary antibiotic binding pockets. TolC forms the conduit and consists of an alpha-helical periplasmic portion and a β-barrel outer membrane crown. AcrA is an adapter protein linking AcrB and TolC within the periplasm. The complex exists in a 3:6:3 stoichiometry for AcrB:AcrA:TolC [20, 21]. Once assembled, the AcrAB-TolC efflux machinery exhibits broad substrate specificity and has been shown to extrude various compounds such as bile salts and detergents [22] to multiple classes of antibiotics such as erythromycin, chloramphenicol, tetracycline, doxorubicin, β-lactams, acriflavine [23, 24]. Efflux pump inhibitors (EPIs) could serve to circumvent this mechanism and work synergistically with current antibiotics to increase their potency.

Because the AcrAB-TolC efflux system requires many linked components to function, there exist multiple avenues to combat its activity. One strategy of exploration is the chemical modification of existing antibiotics [17] so they are no longer substrates for AcrB. However, chemical modification to abolish AcrB binding while retaining the physicochemical properties that afford membrane permeability and inhibitory action on the target can prove to be difficult. Synthetic DNA/RNA mimics that silence mRNA expression of efflux pump components have been shown to increase efficacy of antibiotics [25]. In principle, compounds that uncouple proton motive force from AcrB would collapse the energy required to support drug transport [26, 27]. However, currently known compounds such as carbonyl cyanide m-chlorophenyl hydrazone (CCCP) are not compatible with endogenous mitochondrial function and only serve as experimental chemical tools. Recently, using computational virtual screening methods targeting

the periplasmic adapter protein AcrA, small molecule inhibitors were identified that inhibit mature pump assembly and reduced the minimum inhibitory concentrations for novobiocin and erythromycin [28, 29].

One promising yet under explored tactic is plugging the channels responsible for antibiotic transport. It has been long postulated that an aspartate ring along the interior TolC presents a site for targeting when it was found that it could be blocked by cationic ions [30, 31]. Hexammine cobalt was shown to bind TolC by isothermal titration calorimetry and disrupt channel conductance. However, when determining minimum inhibitor concentrations in combination with erythromycin or fusidic acid it showed no synergistic effect. Targeting the system in this manner however offers the distinct advantage in that the acting agent does not have to cross the double membrane barrier to work on intracellular targets.

I explore a means to target TolC from the extracellular environment by taking inspiration from nature. Colicins are a class of bacterial protein toxins that hijack outer membrane proteins to deliver a cytotoxic domain [32]. In particular, colicin E1 requires two outer membrane protein receptors BtuB and TolC. The N-terminal fragments have been shown to bind to TolC [33, 34]. In chapter 3, I use n-terminal fragments of colicin E1 to see if they can occlude TolC and thus prevent efflux.

Chapter 1

Full and Partial Agonism of a Designed Enzyme Switch†

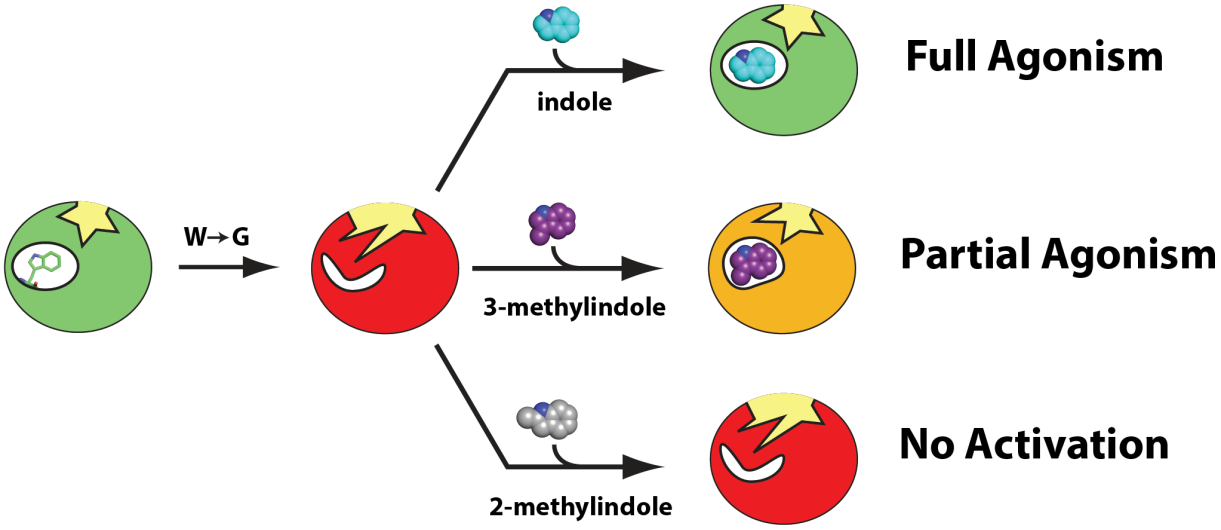
S. Jimmy Budiardjo[1], Timothy J. Licknack[2], Michael B. Cory[2], Dora Kapros[2], Anuradha Roy[3], Scott Lovell[4], Justin Douglas[5], and John Karanicolas[1,2*]

[1]Center for Computational Biology, [2]Department of Molecular Biosciences, [3]High Throughput Screening Laboratory, [4]Protein Structure Laboratory, [5]Molecular Structures Group

The University of Kansas, 2030 Becker Dr., Lawrence, KS 66045-7534

†The text of Chapter 1 is a reprint with permission from: Budiardjo SJ, Licknack TJ, Cory MB, Kapros D, Anuradha R, Lovell S, Douglas J, and Karanicolas J. (2016) Full and Partial Agonism of a Designed Enzyme Switch. *ACS Synthetic Biology*, 5(12), p. 1475-148 Copyright © 2016 American Chemical Society. The supporting information for this chapter is included as Appendix A.1.

Graphical Abstract



Abstract

Chemical biology has long sought to build protein switches for use in molecular diagnostics, imaging, and synthetic biology. The overarching challenge for any type of engineered protein switch is the ability to respond in a selective and predictable manner that caters to the specific environments and timescales needed for the application at hand. We previously described a general method to design switchable proteins, called "chemical rescue of structure", that builds *de novo* allosteric control sites directly into a protein's functional domain. This approach entails first carving out a buried cavity in a protein via mutation, such that the protein's structure is disrupted and activity is lost. An exogenous ligand is subsequently added to substitute for the atoms that were removed by mutation, restoring the protein's structure and thus its activity. Here, we begin to ask what principles dictate such switches' response to different activating ligands. Using a redesigned β-glycosidase enzyme as our model system, we find that the designed effector site is quite malleable and can accommodate both larger and smaller ligands, but that optimal rescue comes only from a ligand that perfectly replaces the deleted atoms. Guided by these principles, we then altered the shape of this cavity by using different cavity-forming mutations, and predicted different ligands that would better complement these new cavities. These findings demonstrate how the protein switch's response can be tuned via small changes to the ligand with respect to the binding cavity, and ultimately enabled us to design an improved switch. We anticipate that these insights will help enable design of future systems that tune other aspects of protein activity, whereby, like evolved protein receptors, remolding the effector site can also adjust additional outputs such as substrate selectivity and activation of downstream signaling pathways.

Introduction

Chemical biology has been used to engineer small molecule-dependent activity into a variety of select proteins: this has allowed external activation of pathways controlling phosphorylation [1], glycosylation [35], and proteolysis [2, 36]. However, each of these advances resulted from implementing pharmacological approaches uniquely tailored to the particular problem at hand. Recently, efforts have expanded to building synthetic switchable proteins in a more general way [3].

The potential utility of generalized protein switches is vast, and already they have successfully been used in certain cases to decipher cellular mechanisms and to construct novel devices such as biosensors [3, 37, 38]. Some of these generalized examples include switches that activate protein function by using a small-molecule to reverse constitutive degradation of the target protein [2, 36] or induce removal of a self-splicing protein element (an "intein") [39-41]. These systems have typically been designed through an engineering-centric "bottom-up" approach [3]: functional modules from different biological systems are mixed and matched to obtain the desired function. In another example, maltose-binding protein (input domain) was fused to β-lactamase (output domain) such that sugar binding regulated β-lactamase activity [42]. However, engineering these switches is still far from rational; it entails extensive trial-and-error and/or directed evolution, particularly with regards to the linkers used for tethering together the functional modules [43]. Individual point mutants can also tune the behavior of such protein switches, through effects on either the input domain (effector binding) and/or the output domain (catalysis) [44]. While combinatorial libraries of randomized gene insertions and their point mutants can be tested to find combinations with the desired effect, it will nonetheless be advantageous to instead design these types of switches in a more rational way: this will enable

the switch's output(s) to not only be predicted, but also tuned for use in different environments and timescales.

As an alternative to this modular engineering strategy, we recently described a method called "chemical rescue of structure" that designs a new small molecule binding site directly into the protein's output domain [5, 6]. We start by identifying a "buttressing" side chain in the protein core (e.g. a tryptophan) that is required to maintain the architecture of the protein functional site; removal of this structural feature by mutation to glycine (e.g. W→G) disrupts this architecture and leads to loss of function. We then restore the buttress by adding the cognate ligand (e.g. indole), which in turn restores the original protein conformation and rescues its activity.

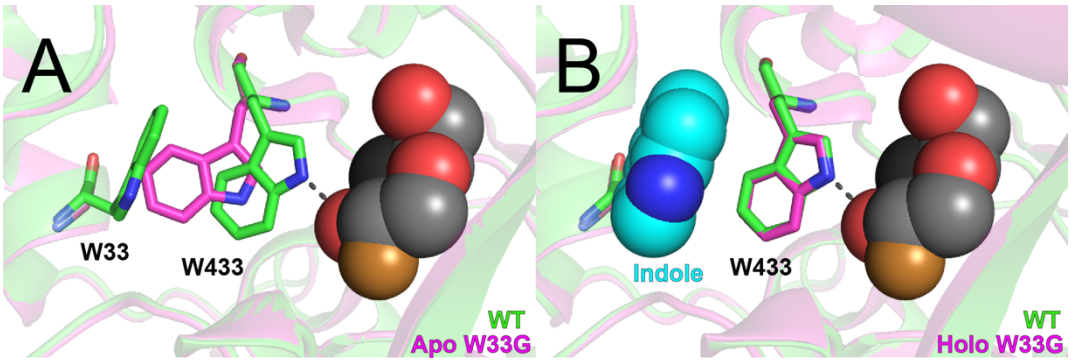


Figure 1.1: Structural basis for inactivation and rescue in β-gly W33G. A comparison of the previously-determined crystal structures of wild type β-glycosidase (*green*) with its W33G mutant (*pink*). A covalent substrate analog (2-fluoro-2-deoxy-D-glucose, *spheres*) is included from WT structure to indicate the location of the active site; this compound was not present in either structure of β-gly W33G. **(A)** In the apo structure of β-gly W33G, W433 shifts into cavity that was previously occupied by W33; thus, W433 is no longer positioned to participate in hydrogen bonding with the substrate. **(B)** Upon soaking with indole, the holo structure of β-gly W33G shows that W433 has returned to its original position; indole (*cyan spheres*) occupies the cavity resulting from the W33G mutation.

Our initial studies used β-glycosidase, a sugar-metabolizing enzyme from *Sulfolobus solfataricus*, as a model system. A high-resolution crystal structure revealed that when we

introduced the W33G mutation, catalytic residue Trp433 shifted back into the newly-formed cavity (**Figure 1.1a**): this explained the observed loss of enzyme activity, because Trp433 engages in key hydrogen bonding interactions with the substrate. Strikingly, by solving a holo crystal structure we found that indole binding shifts Trp433 back to its former position in the wild type conformation (**Figure 1.1b**): this explained the observation that addition of indole restores enzyme activity. Using the Michaelis-Menten model to interpret the enzyme kinetics, we found that introduction of the W33G mutation led to a 730-fold decrease in k_{cat}/K_m relative to wild type; this was completely restored upon addition of 10 mM indole [5].

Whether from "chemical rescue of structure" or from some other strategy, a current limitation in designing new molecular switches is the ability to selectively tune the responses of the switches in a predictable manner, as compared to the many other fields of engineering in which very precise control can be achieved [45]. Biological systems are inherently noisy and complex, making it difficult to predict in a rational way how a given system will react to a particular input signal [46]. In natural systems, such as G protein-coupled receptors (GPCRs), receptors have evolved to interact with a wide range of regulatory molecules [47] and can interact with different ligands that induce opposing responses [48]. In estrogen receptor-α, ligand binding dynamics can dictate the fate of unique signaling pathways [49]. In extreme cases, such as certain transcription factors, a single ligand can induce opposing responses on a given protein (agonist versus antagonist), depending on the context [50].

With regards to "chemical rescue of structure," this comparison to natural signaling proteins prompts analogous questions. First, how should a cavity forming mutation be designed and then complemented in order to achieve optimal rescue? Second, how malleable is the resulting effector-binding site? In the case of the β-glycosidase example, the indole concentration

can be varied in order to tune V_{max}. But natural systems are rarely regulated by just one ligand – to what extent will this synthetic system respond to other ligands?

We envision two potential underlying structural mechanisms for rescue. In the first model the effector site behaves as a discrete switch, such that ligand binding exactly restores the geometry of the active site; such behavior is precedented by a recent study showing the discrete response of T4 lysozyme to a series of congeneric ligands [51]. Given such a model, the degree to which a given ligand activates the protein would be related solely to the ligand's binding affinity at the effector site. In the second model, the effector site responds in a continuous fashion: more diverse ligands can be accommodated, and each may influence protein structure – and thus activity – slightly differently. In this model, activation by a ligand is driven not just by its binding affinity, but also by the extent to which the ligand precisely restores the active site geometry for catalysis; analogous behavior is observed in crystal structures of GPCRs bound to full and partial agonists, in that the partial agonists make only a subset of the specific interactions of the full agonists [52].

In this study we seek to distinguish between these two models, and to further define the ideal chemical and geometric characteristics of a maximally rescuing ligand. We anticipate that the governing properties of this system will shed insight on the tunability of molecular switches built from "chemical rescue of structure" for future applications.

Results

Structure-activity relationship of alternate effector ligands

To probe the malleability of the effector site, we began by exploring the structure-activity relationship of rescue using a focused library of fifteen indole analogs with the β-glycosidase

W33G enzyme switch described above. To probe the effector site cavity (the void left behind by the Trp to Gly mutation), we compiled this set by adding or removing functional groups around the indole parent scaffold. In order to explore in-plane substituents, methyl groups were individually added to each of the seven outer positions. To probe the effect of removing the hydrogen bond donor and allowing ring pucker, we tested indene and indan. We also included naphthalene and quinoline to explore the effect of increasing the ring size slightly, and we included benzene, toluene, and N-methylaniline to explore the effect of smaller compounds.

We tested each of these compounds in a β-glycosidase enzyme assay, to determine the extent to which each compound rescued activity (**Figure 1.2**). We used a high substrate concentration relative to the rescued enzymes' K_m values (**Figure A1.1**), such that our assay was designed to provide a readout on the effector's influence on k_{cat}. Because some effector compounds inhibited the wild type enzyme slightly at this high concentration, we normalized all results to the wild type enzyme under the same conditions. Each compound was tested at a concentration of 2 mM, because this was the maximum concentration at which all 15 compounds were soluble in our assay conditions.

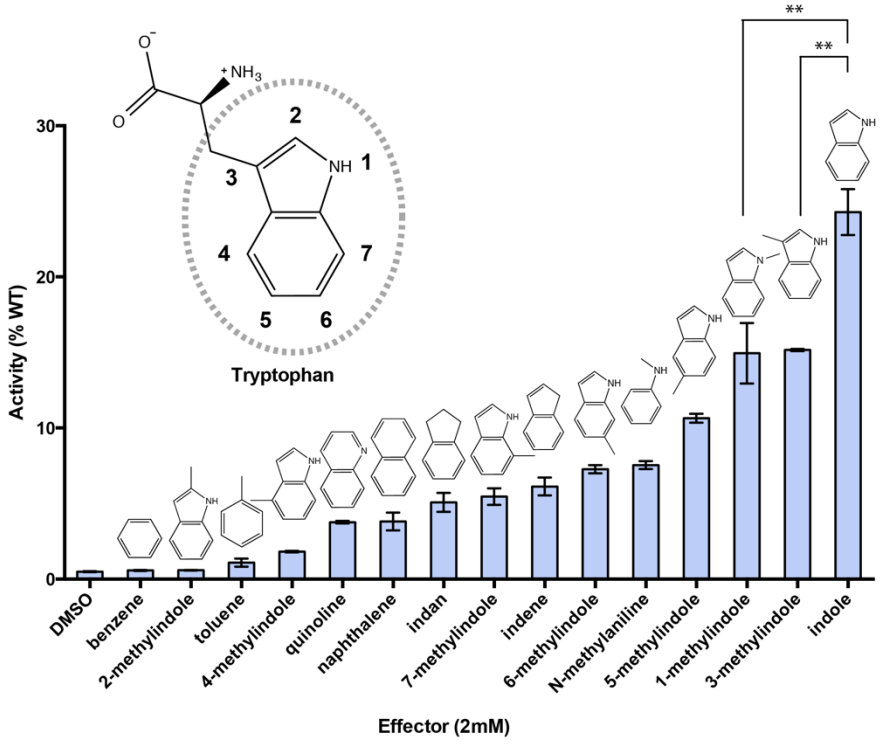


Figure 1.2: Probing malleability of the β-gly W33G effector site. The initial velocity is shown relative to the wild type enzyme in the presence of the same effector ligand; compounds are presented in ascending order of activity. Product formation was measured spectrofluorometrically using 750 μM fluorescein di-β-D-galactopyranoside (FDGal) as a substrate; each effector ligand was tested at 2 mM. All data were collected in triplicate, and error bars correspond to the standard error of the mean. Bars marked with ** have a statistically significant difference in their mean values ($p < 0.01$).

Relative to the wild type enzyme, W33G retains 0.5% basal activity in the absence of any ligand. In our previous study we found that indole fully restored wild type activity at a concentration of 10 mM [5]; here, indole only restores 25% of the wild type activity because a lower concentration is used (2 mM). Of the 15 compounds tested, we found that 13 of them rescued activity; however, the extent of rescue was quite diverse. Among the seven methylindoles, for example, we observe the most rescue from 1- and 3-methylindoles; in contrast, 2-methylindole yields no increase in activity over the apo enzyme. Collectively these results show that structural details of the effector ligands can produce differing levels of catalytic activity – either through differences in binding affinity, or through conformational differences in

the holo enzyme. In order to distinguish these possibilities, we next sought to directly detect binding of these ligands to the enzyme.

Direct determination of effector binding

We previously estimated the binding affinity of indole to β-gly W33G to be 0.75 mM in the presence of saturating substrate, and 15 mM in the absence of substrate [5]. The weak binding affinity of this interaction – and presumably those of the indole analogs – places it outside the sensitivity limits of many common approaches for detecting protein-ligand interactions. To overcome this hurdle we therefore turned to ^{19}F NMR [53, 54]: the large chemical shift anisotropy of ^{19}F nuclei results in line width differences of the fluorine signal between its free and protein-bound states, which is highly amenable for detection of weak interactions [55]. In other words, differences between the transverse relaxation times of the free and protein-bound states are manifest through broadening of the ^{19}F signal [55, 56].

We began by identifying a fluorine-containing reporter molecule: we selected 6-fluoroindole, having found that this compound rescues enzyme activity at a very similar level as indole itself (**Figure A1.2**). Free in solution, the fluorine in this compound exhibits a sharp peak at -122.07 ppm. Upon addition of WT β-gly there is no change to this peak; however, upon addition of β-gly W33G we observe considerable peak broadening indicating that the reporter binds to the cavity-containing mutant (**Figure 1.3a, Table A1.1**). Importantly, for this experiment we included 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside, a covalently-attached substrate analog of this enzyme: given the allosteric linkage between the active site and effector site, we reasoned that inclusion of a substrate analog would order the active site residues for catalysis, and in doing so would also pre-order the effector site residues for ligand binding.

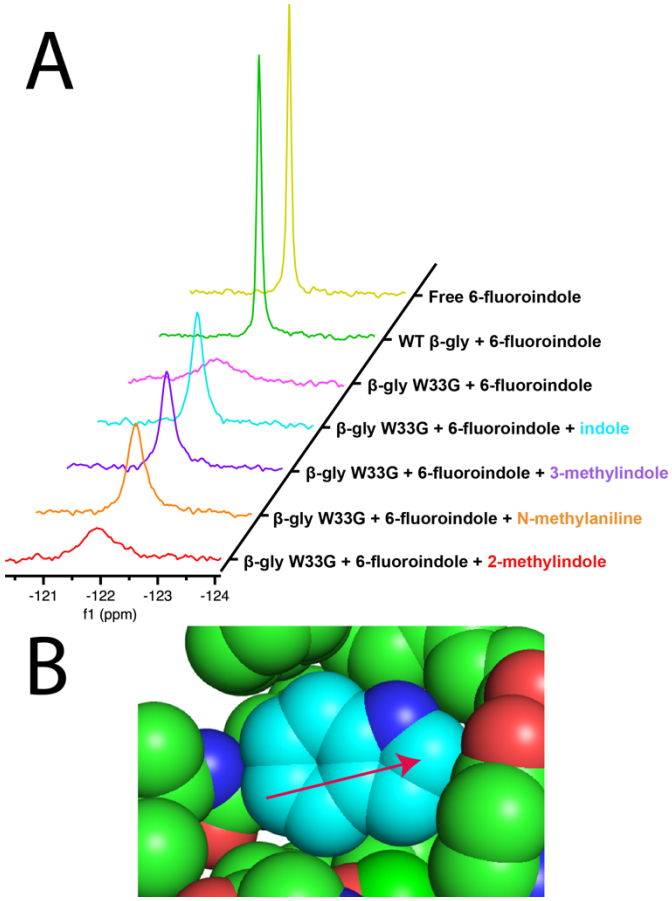


Figure 1.3: Direct determination of effector binding using a ^{19}F NMR competition assay. **(A)** 6-fluoroindole (2 mM) is used as a reporter molecule in all cases, with 500 μM protein concentration. Each competitor ligand was tested at a concentration of 5 mM. The sharp peak observed for the unbound reporter is broadened upon binding to the protein; ligands that compete for this binding site displace the reporter, leading to peak sharpening. **(B)** The packing of wild type β-glycosidase around Trp33 is tightest around the 2-position (*arrow*), making it particularly difficult to accommodate an extra methyl group here.

Next, we evaluated the ability of several ligands to compete with 6-fluoroindole for binding to β-gly W33G (presumably at the allosteric effector site). In addition to indole, we selected 3-methylindole (the next best rescuing compound), 2-methylindole (no rescue), and N-methylaniline (a smaller compound with intermediate rescue).

Upon addition of indole to our sample, we find that the ^{19}F signal sharpens (**Figure 1.3a, Table A1.1**): this implies that indole displaces some of the 6-fluoroindole from the effector site.

We also observe analogous peak sharpening for both 3-methylindole and N-methylaniline. In contrast, addition of 2-methylindole produces very little, if any, peak sharpening; this suggests that 2-methylindole binds to β-gly W33G much less strongly than 6-fluoroindole, and the other analogs included in this experiment.

This in turn provides an explanation for why 2-methylindole rescued activity less than other analogs (**Figure 1.2**). It is *not* the case that 2-methylindole binds to the allosteric effector site in a manner unproductive for catalysis; rather, 2-methylindole fails to rescue activity because it simply does not bind to this site. This explanation is further consistent with the protein structure (**Figure 1.3b**): the protein environment around Trp33 in the wild type enzyme is tightest around the 2-position, making this the most sterically challenging location at which to accommodate an extra methyl group.

^{19}F NMR competition assays have been used to measure dissociation constants ranging from low nM to high μM [56], but unfortunately this experiment does not allow us to determine the binding affinities of these ligands in an accurate and quantitative way. Nonetheless, as a qualitative measure of binding this assay complements the enzymatic functional assay that will be presented below, for discriminating non-rescuing compounds that do not bind, versus those that bind but do not rescue.

Effect on rescue of both the substrate and the effector ligand

To directly examine potential differences in the active site geometry resulting from rescue by alternate ligands, we soaked crystals of β-gly W33G with 3-methylindole, with N-methylaniline, and with several other ligands. While soaking with indole led to a very useful crystal structure of that complex [5], upon soaking with other ligands we found only the unbound

protein. Attempts to co-crystallize other ligands with β-gly W33G also yielded the unbound protein. For this reason, we sought to probe for small structural differences *indirectly*, through effects on substrate recognition.

Others have shown for allosteric activators of an unrelated system that the level of activation can depend on the structural properties of a specific substrate, and is thus a function of the precise arrangement of the activator-enzyme-substrate complex [57]. Our previous study implicated Trp433 as the switch that distinguishes the active and inactive states of β-gly W33G. Accordingly, we anticipated that the precise placement of Trp433 in response to the rescuing ligand would be crucial for dictating the level of catalysis in the holo enzyme – and that rescue may be dependent not only on the effector ligand, but also on the choice of substrate.

The structure of the wild type β-glycosidase has been solved in complex with two different covalent substrate analogs, 2-fluoro-2-deoxy-D-galactose (2F-Gal) and 2-fluoro-2-deoxy-D-glucose (2F-Glc) [58]. While these two inhibitors form nearly identical interactions with the active site, the primary difference in these two structures is the interaction of the substrate analog with our "switch" residue, Trp433: 2F-Gal requires that the Trp433 sidechain splits its hydrogen bonding potential between the C3 and C4 hydroxyls (**Figure 1.4a**), whereas the different stereochemistry at the C4 position in 2F-Glc allows the Trp433 sidechain to hydrogen bond only with the C3 hydroxyl (**Figure 1.4b**). Based on this difference we anticipated that Gal-derived substrates would be more sensitive to structural perturbations of Trp433 than Glc-derived substrates. As a consequence, we further anticipated that Glc-derived substrates would be easier to rescue by effectors that do not perfectly fit the cavity, and may therefore not restore the perfectly ideal catalytic geometry of Trp433.

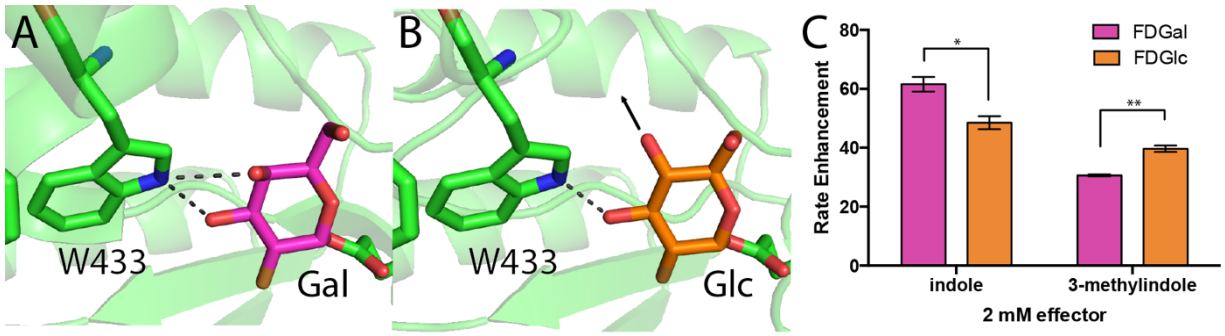


Figure 1.4: Trp433 forms different interactions with Gal versus Glc, allowing effectors to rescue to different extents, leading to slightly different effector preferences. Crystal structures of wild type β-glycosidase bound to covalent substrate analogs **(A)** 2F-Gal and **(B)** 2F-Glc. The active site residue Trp433 makes distinct hydrogen bond interactions depending on the sugar's stereochemistry at the C4 position. **(C)** Enhancement of initial velocity (relative to the apo enzyme) for substrates that differ only at this stereocenter (FDGal versus FDGlc), rescued using either indole or 3-methylindole. Assays were carried out using 250 μM substrate and 2 mM effector. Bars indicated have a statistically significant difference in their mean values (* with $p < 0.05$, ** with $p < 0.01$).

To test this hypothesis, we compared rescue using Gal-derived substrates versus Glc-derived substrates. Starting with the same Gal-derived substrate utilized in **Figure 1.2** (fluorescein di-β-D-galactopyranoside, FDGal), we again observe better rescue with 2 mM indole than with 2 mM 3-methylindole. However, when we instead use the analogous Glc-derived substrate (fluorescein di-β-D-glucopyranoside, FDGlc) the difference in rescue between these substrates is greatly reduced (**Figure 1.4c**). As anticipated, then, rescue of the Gal-derived substrate depends more on precisely recapitulating the catalytic geometry of Trp433. If its position were perfectly superposed with the Trp33 sidechain in the wild type enzyme, the additional methyl group in 3-methylindole would align to the original Cβ position; in the rescued mutant, this would form a steric clash with the Cα to which it was previously covalently attached. We therefore expect that the 3-methylindole location would be slightly shifted relative to Trp33 (and relative to rescuing indole), which in turn explains its slightly decreased ability to rescue activity against FDGal. Though subtle, the ability of these rescuing ligands to shift the

substrate preference for this enzyme is also highly reminiscent of "functional selectivity" in the pharmacology of natural signaling systems [59-62].

Separating effects of binding affinity from effects of active site geometry

To definitively separate effects of binding affinity from effects on the active site geometry, we next examined the concentration dependences of these effectors on the rate of product formation, using FDGal as substrate. To mitigate potential effects on K_m (and thus simplify analysis), we again carried out experiments in a regime at which the substrate concentration (750 μM) is higher than K_m of the rescued enzyme (**Figure A1.1**); thus, the observed initial velocity depends primarily on V_{max} (which, in turn, depends in part on the effector concentration).

To fit the resulting initial velocities, we adapted the rate equation of a simple allosteric kinetic mechanism [63] by simplifying under the limit of saturating substrate (see *Methods*). This gave an equation that relates the change in initial velocities due to rescue to two key parameters, K_D and W. K_D is the effector dissociation constant, and W is a "linkage" term describing the magnitude of the effect of the allosteric ligand on V_{max}. Structurally, the term W can also be interpreted as the extent to which the rescuing ligand restores the enzyme's catalytic activity, because it reports on the rate of the holo enzyme.

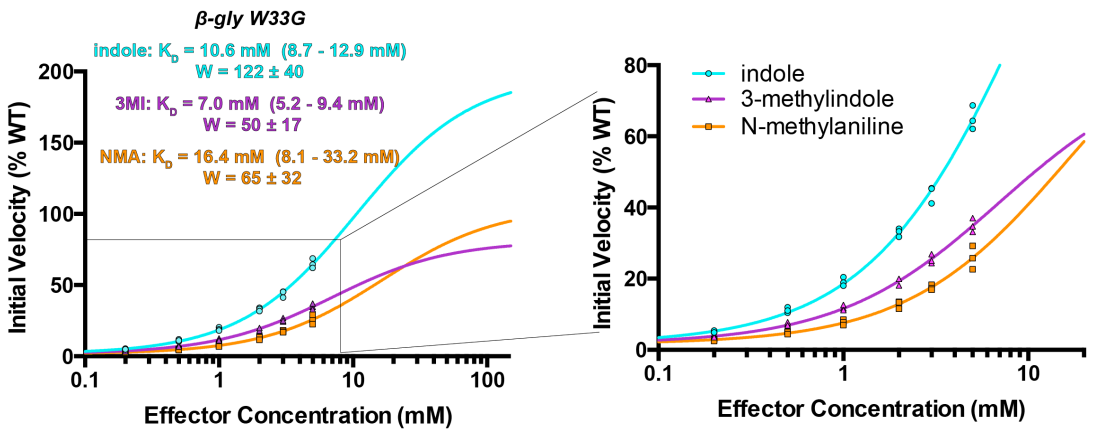


Figure 1.5: Full and partial agonism of β-gly W33G. Enzyme assays carried out at high substrate concentration (750 μM FDGal) demonstrate the dependence of the initial velocity (and thus V_{max}) on effector concentration. Both panels show the same data, expanded in the right panel to better show the agreement between the data and the model that was fit to it. All data were collected in triplicate, with all replicates shown explicitly on the plots. Experimental uncertainly of the fit values for K_D and W are reported as a range corresponding to one standard error in either direction.

Indole, 3-methylindole, and N-methylaniline each produce dose-dependent increases in the initial velocity that are well-described by this equation (**Figure 1.5**). To remain within the solubility limits for all three compounds, we were constrained to a maximum concentration of 5 mM. The curves have not fully saturated by this concentration, and so we present two views of the data: one that includes the extrapolated curve fitting (to show the full behavior of the model), and one that instead focuses only on the region for which experimental data was obtained (to show the agreement of the data to the model). Through these curve fits, the values of K_D and W are obtained in tandem. To clearly establish the limits of this extrapolation, we report the bounds of K_D and W (one standard error in either direction); the resulting bounds on the fitted curves are shown visually as well for comparison (**Figure A1.3**).

Among these three compounds, indole has the highest linkage term (W); this is consistent with our expectation that indole most effectively rescues the precise geometry of the active site.

In fact, our extrapolation suggests that at high enough concentration the indole-rescued W33G mutant would actually surpass the rate of the wild type enzyme. While we cannot quite access these concentrations due to indole's solubility, it is possible that the additional flexibility of the rescued enzyme allows W433G to adopt a conformation very slightly different than that of the wild type enzyme (and even more favorable for catalysis).

While the linkage term (W) for 3-methylindole is unsurprisingly worse than for indole, the K_D value is actually slightly better than for indole. Thus, 3-methylindole appears to bind to β-gly W33G at least as tightly as indole, but does not rescue activity to the same extent because it does not fully restore the catalytic geometry. This observation is also consistent with those from the experiments described in the previous sections.

Finally, we compare the behavior of N-methylaniline to that of 3-methylindole. While the fitted values for W and K_D are very similar for both ligands, the small differences provide an opportunity to visually highlight how these parameters together dictate the response to an effector ligand. 3-methylindole has a superior K_D and thus more activity at low effector concentrations; in contrast, N-methylaniline has a superior W value and thus more activity at high effector concentrations – albeit at concentrations accessible only via extrapolation, due to solubility restrictions.

Drawing from long-established "receptor theory" pharmacology of natural signaling systems [64], 3-methylindole and N-methylaniline behave as *partial agonists* for the β-gly W33G "receptor", in relation to the *full agonist*, indole.

Designing a better activating ligand for β-gly W33G

To test this model of β-gly W33G activation, we sought to rationally design a ligand with enhanced signaling relative to indole itself. Given our model, it will be very difficult to design a compound with more activity at high effector concentrations (better W), since this would require shifting the catalytic residues in a very prescribed way. Instead, we focused on ligands that might have better activity than indole at low effector concentrations, where binding affinity (better K_D) is most responsible for determining activity.

Selectively adding fluorine is a common step in medicinal chemistry when seeking to improve potency of a compound: this makes the compound more lipophilic and also allows for productive halogen-pi interactions [65, 66]. Because fluorine is isosteric with hydrogen, shape complementary for the receptor is typically preserved. Further, fluorine-substituted indoles (such as the reporter compound used in the ^{19}F NMR competition assay) at almost all positions are commercially available.

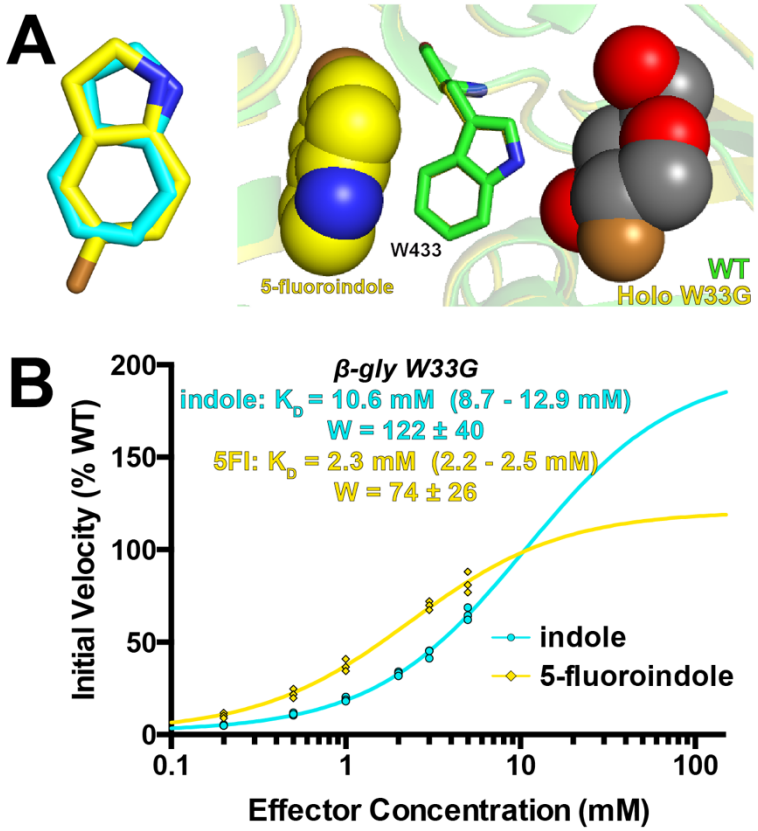


Figure 1.6: 5-fluoroindole rescues β-gly W33G with a different signaling profile than indole. (A) The crystal structure of 5-fluoroindole, solved to 1.75 Å resolution. At *left*, superposition with indole shows that 5-fluoroindole occupies essentially the same binding model (electron density for 5-fluoroindole is presented in **Figure S4**). At *right*, superposition with the wild type β-glycosidase shows that 5-fluoroindole restores the position of Trp433 (superposition with the indole-bound structure is presented in **Figure S5**). **(B)** β-gly W33G activity can be rescued either by indole or 5-fluoroindole. 5-fluoroindole binds more tightly, and thus rescues more effectively at low concentrations. At higher concentrations the difference in activity is reduced, and extrapolation suggests that indole will provide more activity than 5-fluoroindole at saturation. In all cases 750 μM FDGal was used as substrate. All data were collected in triplicate, with all replicates shown explicitly on the plots. Experimental uncertainly of the fit values for K_D and W are reported as a range corresponding to one standard error in either direction.

To begin, we solved the crystal structure of one of these, 5-fluoroindole, in complex with β-gly W33G (**Figure 1.6a**). Unsurprisingly, we found that 5-fluoroindole engages β-gly W33G in the same manner as indole, with no notable differences relative to the indole-bound or wild type structures of the enzyme. We then characterized rescue of activity from 5-fluoroindole, as

we previously did for 3-methylindole and N-methylaniline. Indeed, relative to indole we find that 5-fluoroindole provides enhanced β-glycosidase activity at each of the concentrations we tested (**Figure 1.6b**). Notably, however, this difference is manifest most at low effector concentrations; as the concentration is increased indole starts to catch up with 5-fluoroindole, and is extrapolated to ultimately surpass the activity induced by 5-fluoroindole. This behavior is captured quantitatively through the parameters of each fit, in which K_D favors 5-fluoroindole but W favors indole. While the crystal structure cannot provide an explanation for the difference in W, we attribute this to a very small difference in either the structure or dynamics of Trp433 in response to these two ligands.

Designing activating ligands for β-gly W33A

Encouraged by the ability to rationally design an altered response by varying the rescuing ligand, we next sought to adjust the shape of the β-gly cavity, and test whether complementary rescuing ligands could be designed to these new shapes. As a first experiment, we reduced the size of the cavity by incorporating W33A instead of W33G. Earlier, we explained the preference of β-gly W33G for indole over 3-methylindole (**Figure 1.5**) by pointing out that while 3-methylindole restores the same number of removed (non-hydrogen) atoms that were removed by mutation, rescue would require close contact between non-covalent atoms that were previously bonded to one another (Cα and Cβ). Applying this logic to W33A, then, one would expect a steric clash between the Cβ of Ala33 and the atom replacing the Cγ of the original tryptophan sidechain (i.e. the indole 3-position). Instead, one would predict better rescue by the corresponding compound that lacks this atom: N-methylaniline. Spurred by this rationale, we therefore explored rescue of β-gly W33A by N-methylaniline, indole, and 3-methylindole.

Consistent with this expectation, we find that N-methylaniline rescues β-gly W33A better than either indole or 3-methylindole (**Figure 1.7**); however, the difference was surprisingly modest. A further surprise was the fact that 3-methylindole rescued activity better than indole, which was counterintuitive given that this experiment entailed reducing the size of the cavity.

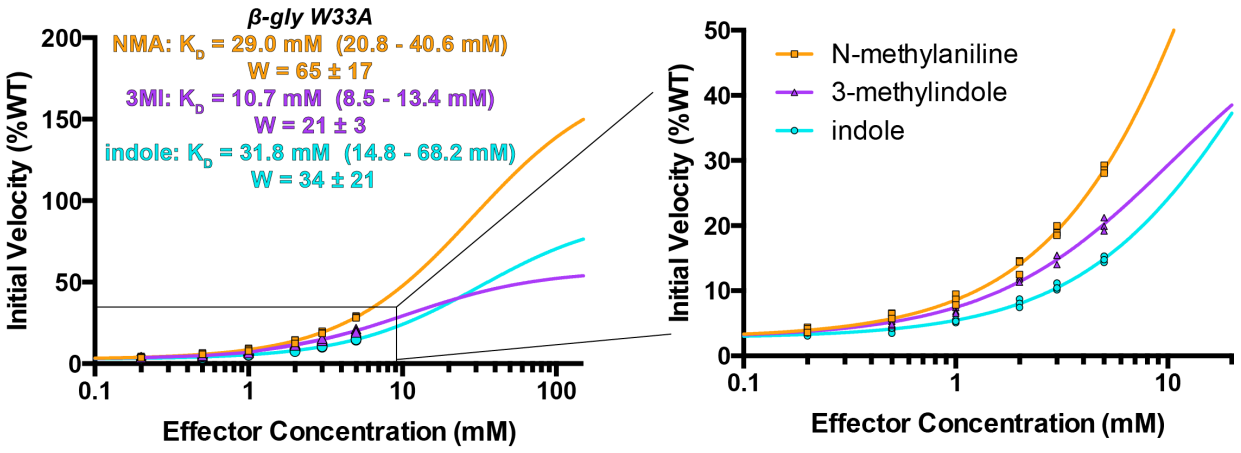


Figure 1.7: N-methylaniline is the optimal rescuing ligand for β-gly W33A. The relatively modest rescue by N-methylaniline, as well as the fact that 3-methylindole rescued better than indole, can be explained by separately examining the binding affinity (K_D) and contribution to the active site geometry (W) for each compound. In all cases 750 μM FDGal was used as substrate. All data were collected in triplicate, with all replicates shown explicitly on the plots. Experimental uncertainly of the fit values for K_D and W are reported as a range corresponding to one standard error in either direction.

The explanation for both of these puzzles is found in the separate contributions from active site geometry (W) and binding affinity (K_D). N-methylaniline was expected to be most compatible with the Trp433 geometry that is optimal for catalysis; indeed, N-methylaniline has the highest value of W. Its relatively modest activation, meanwhile, derives solely from its poor binding affinity, which may reflect the entropic cost of ordering its freely rotatable bond (that is not present in the other compounds). In the same vein, the surprisingly good activity of 3-methylindole relative to indole also derives exclusively from a difference in binding affinity, presumably because 3-methylindole is more hydrophobic than indole. While it remains surprising that the value of W is not much worse for 3-methylindole than for indole, we speculate

that 3-methylindole may flip to position its methyl group towards the exposed region used by indole's amine group (**Figure 1.3b**); this hypothesis could best be tested through crystallographic evidence, however we have been unable to solve a structure of either β-gly mutant in complex with 3-methylindole. Finally, we note that the 3-methylindole and indole values of *W* are lower for W33A than for W33G; this is also consistent with the hypothesis that neither of the compounds can restore the active site structure quite as accurately when disrupted by Ala33.

In summary, while N-methylaniline proved to be the optimal ligand for rescuing this smaller cavity, the magnitude to which catalytic activity could be restored was nonetheless limited by the small size and flexible nature of this compound. To design a more sensitive switch, we therefore introduced a larger cavity into this enzyme.

Designing activating ligands for β-gly W33G_V37A

Based on the structure of the wild type enzyme, there was only a single nearby sidechain in plane with the indole ring of Trp33, Val37 (**Figure 1.8a**). To expand the allosteric effector site, we therefore introduced an additional mutation, V37A, into β-gly W33G. Based on the location of this additional cavity-forming mutation, we expected that β-gly W33G_V37A would best be rescued by indole substituted at the 5-position.

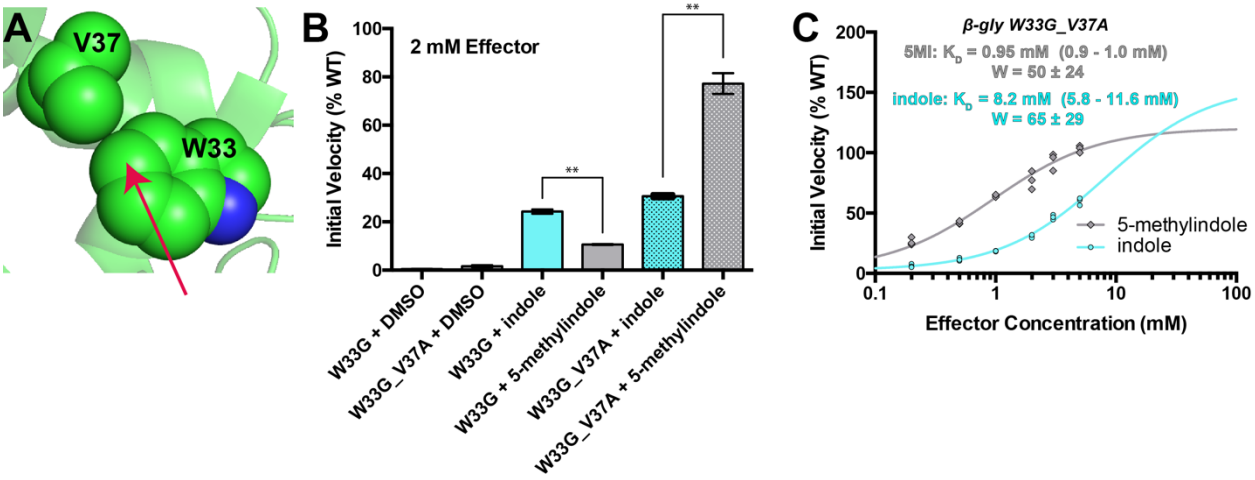


Figure 1.8: 5-methylindole effectively rescues β-gly W33G_V37A. (A) Based on the crystal structure of the wild type enzyme, Val37 was a natural location at which to expand the allosteric effector site. We anticipated that reducing the size of this sidechain would allow an additional substituent at indole's 5-position (*arrow*). **(B)** At a concentration of 2 mM, indole rescues the β-gly W33G and W33G_V37A mutants about equally well. Relative to indole, 5-methylindole rescues β-gly W33G much *less*, and β-gly W33G_V37A much *more*. Error bars correspond to the standard error of the mean, and bars marked with ** have a statistically significant difference in their mean values ($p < 0.01$). **(C)** By separately examining the binding affinity (K_D) and contribution to the active site geometry (W), it is clear that the basis for the difference in rescue of W33G_V37A derives from the tighter binding affinity of 5-methylindole for this double mutant. In all cases 750 μM FDGal was used as substrate. All data were collected in triplicate, with all replicates shown explicitly on the plots. Experimental uncertainly of the fit values for K_D and W are reported as a range corresponding to one standard error in either direction.

We initially tested the activity of both indole and 5-methylindole against β-gly W33G, at a concentration of 2 mM (**Figure 1.8b**). As in our previous experiment (**Figure 1.2**), we found that indole conferred about twice the activity of 5-methylindole under these conditions. Testing these two compounds against β-gly W33G_V37A, however, showed a reversal of their activities: now 5-methylindole proved superior over indole.

To understand the basis for this difference, we explored β-gly W33G_V37A activity as a function of effector concentration (**Figure 1.8c**). Perhaps unsurprisingly, both compounds have very similar values of W: this implies that both compounds engage the effector site with very similar binding modes, and thus interact with Trp433 in the same way. Accordingly, the basis for

the greater potency of 5-methylindole comes from its binding affinity (K_D), which is almost an order of magnitude tighter than that of indole. This observation is also unsurprising, given the extra hydrophobic surface area that is buried upon binding of 5-methylindole instead of indole.

At the outset of this study, we tested a panel of indole analogs at 2 mM for rescue of β-gly W33G, and found that none of these surpassed the 25% of wild type activity recovered by indole itself (**Figure 1.2**). Now, with a more refined view of this system predicated on pharmacological receptor theory [64], we have designed a new protein-ligand pair that exhibits almost 80% of wild type activity at 2 mM effector (**Figure 1.8b**).

Discussion

In the introduction to this study, we laid out two potential structural mechanisms for rescue at this designed allosteric effector site. In the first model the effector site behaves as a discrete active/inactive switch, activated through ligand binding. In the second model the effector site behaves as a rheostat, reporting not only on the presence of an activating ligand, but also on its precise structural complementarity for the effector site.

All of the accumulated evidence in our study points to the latter model. Numerous methylindoles show activity (all but 2-methylindole), highlighting the malleability of this site for accommodating alternate ligands. While certain ligands bind to this site more tightly than indole, none rescue the precise active site geometry optimal for catalysis quite as effectively as indole: instead, they serve as partial agonists of β-gly W33G.

Through the structure-activity relationship compiled for β-gly W33G, two principles emerge that guide design of agonists to fit a given cavity. First, the ideal rescuing ligand should replace the protein atoms removed by mutation as closely as possible: this leads to the most

activity once the ligand is bound. Second, there may be a small tradeoff between a steric clash and a small cavity when replacing the deleted atoms: as demonstrated through indole's rescue of β-gly W33G and N-methylaniline's rescue of β-gly W33A, the small unfilled volume is less detrimental for activity than the structural rearrangement required to resolve the steric clash. With these design principles in mind, it became possible to rationalize the complete behavior of many different activating ligands, and also to design a new switch with enhanced activity.

The key strategy motivating design of the 5-methylindole / β-gly W33G_V37A switch was to improve binding affinity while at the same time maintaining the precise complementarity that restored the optimal geometry of Trp433. Given the very small ligands (indole has only 9 non-hydrogen atoms), there is an intrinsic biophysical limit to the potential binding affinity that can be achieved [4, 67]. Indeed, the ligand efficiency (binding free energy per non-hydrogen atom) [67] of indole binding to β-gly W33G is -0.31 kcal/mol·atom (**Table A1.2**), which is comparable to the median ligand efficiency of other protein-ligand complexes (-0.34 kcal/mol·atom) [4]. By design, 5-methylindole incorporates an additional potent interaction (burial of the methyl group in a tight cavity), increasing the ligand efficiency to -0.43 kcal/mol·atom. Given the typical limits of ligand efficiency [4], designing switches that respond to lower ligand concentrations will likely require the use of ligands that rescue larger cavities, or more sophisticated mechanisms that incorporate feed-forward signaling.

In our studies of other model systems activated by indole after incorporation of W→G cavity-forming mutations, we found that inactivation and rescue could be mediated by local or global unfolding/refolding, rather than through a discrete conformational change [6]. This mechanism enabled the use of effector sites located much further from the active site, but may come at the expense of sensitivity towards slightly different ligands. Here, we demonstrated that

the response to different ligands could be tuned through a combination of binding affinity (K_D) and precise structural complementarity (W). We anticipate that longer-ranged allosteric switches built in manner are less likely to preserve the subtle structural response to different ligands as the signal is transduced through the protein; as a result, the response may instead be dominated simply by the binding affinity. The same limitation is also likely to switches that are built by recombining modular domains [45, 68], because it is difficult to predict precisely how the linkers will transduce the input signal over to the output domain. An important advantage of chemical rescue of structure is the ability to place the location of the effector site in close proximity to the output signal (i.e. the active site), which in turn may facilitate the graded responses we demonstrate here.

Looking forward, careful modulation of the response over a short distance may also facilitate design of switches that exhibit functional selectivity. Already we observe this property in β-gly W33G – though not explicitly designed – in the fact that indole and 3-methylindole tune the enzyme's preference for substrates containing galactose versus glucose. Natural systems make extensive use of this paradigm, as exemplified by GPCRs that initiate signaling through either the canonical G protein-mediated pathway or the non-canonical β-arrestin G protein-independent pathway, depending on the particular ligand that it bound [48, 69, 70]. We anticipate that the design principles emerging from this study of the β-gly W33G designed allosteric effector site will provide a first step towards rationally designing new synthetic switches capable of absorbing information from several different molecular cues, and providing distinct and meaningful responses to each of them [3].

Methods

A complete description of methods is available as Supporting Online Materials. Coordinates and structure factors for the crystal structure of β-gly W33G bound to 5-fluoroindole has been deposited with the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) with accession code 5IXE.

^{19}F NMR competition assay

The weak affinity of the interactions we sought to measure places them outside the sensitivity limits of techniques to such as surface plasmon resonance (SPR), isothermal titration calorimetry (ITC), and differential scanning fluorimetry (DSF). The sensitivity problem is exacerbated for some of these techniques by the particularly small ligands (< 200 Da) and the relatively large protein (56,000 Da). We recognized that the problem of detecting very weak binding of a small ligand is reminiscent of the challenges faced in fragment-based drug discovery campaigns, and therefore borrowed an emerging tool from their repertoire, ^{19}F NMR [53, 54].

^{19}F NMR spectra were acquired on a Bruker DRX spectrometer equipped with an 11.7 T magnet (^{19}F resonance frequency equals 470 MHz). 500 μM β-glycosidase in 50 mM phosphate buffer in H_2O with 10% protonated DMSO, 2 mM 6-fluoroindole and 5 mM of the competitor ligand.

Protein samples were pre-treated with 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside. The 2,4-dinitrophenol serves as a leaving group such that the 2-deoxy-2-fluoro-β-D-glucose remains covalently attached to the protein. We confirmed labeling of the protein by broadening of the inhibitor's ^{19}F NMR peak, and also spectrophotometrically by production of 2,4-dinitrophenol.

β-glycosidase enzyme assay

All β-glycosidase enzyme assays were conducted using 58 nM β-glycosidase with fluorescein di-β-D-galactopyranoside (FDGal) or fluorescein di-β-D-glucopyranoside (FDGlc) as substrate, in a buffer of 50 mM sodium phosphate pH 6.5 and 10% DMSO at 37°C. Upon catalysis, FDGal is cleaved twice yielding one D-galactose and two molecules of fluorescein. We detect product formation by following fluorescence with excitation at 485 nm and emission at 528 nm.

Rate equation

To fit the observed initial velocities, we adapted the rate equation of a simple allosteric kinetic mechanism [63] by simplifying under the limit of saturating substrate. This gave the relationship:

$$V = V^{apo} + \frac{V^{apo}\,(W-1)\,[A]}{K_D+[A]} \qquad \text{(Equation 1.1)}$$

where V is the initial velocity at a given effector concentration, [A] is the effector concentration, K_D is the effector dissociation constant, V^{apo} is the maximal velocity in the absence of effector, and W is a "linkage" term describing the magnitude of the effect of the allosteric ligand on V_{max}. Functionally, W is defined as the ratio of the maximal velocity at saturating effector, to the maximal velocity in the absence of effector (in other words, W is the fold-increase in V_{max} that the effector can bring about). Throughout this work we interpret W as the extent to which the rescuing ligand restores the protein structure required for catalysis, because it reports on the rate of the holo enzyme.

Statistical Analysis

Statistical significance of differences in the standard error of the mean were evaluated using two-tailed t-tests (**Figures 1.2, 1.4, and 1.8**).

Standard errors for log(K_D) and W were calculated with GraphPad Prism version 6.0 via:

$$SE(P_i) = \sqrt{\left({SS}/{DF}\right) * (Cov(i,i))} \qquad \text{(Equation 1.2)}$$

where P_i represents the ith adjustable (non-constant) parameter. SS is the sum of the squared residuals. DF is the degrees of freedom (the difference between the number of data points and parameters fit by regression). $Cov(i,i)$ is the i-th diagonal element of the covariance matrix.

The standard error of log(K_D) was used to calculate the upper and lower bounds of the K_D values (one standard error above/below the best-fit value).

Associated Content

A complete description of experimental methods and procedures. **Figure A1.1** showing K_m values for W33G when rescued by three different effectors. **Figure A1.2** showing indole and 6-fluoroindole activity for β-gly W33G. **Figure A1.3** showing upper and lower bounds of initial velocity at ± 1 standard error. **Figure A1.4** showing electron density maps of 5-fluoroindole bound to β-gly W33G. **Figure A1.5** showing superposition of 5-fluoroindole-bound and indole-bound crystal structures. **Table A1.1** containing ^{19}F NMR peak integrals. **Table A1.2** containing ligand efficiency values for each ligand/protein pair reported in this study. **Table A1.3** containing crystallographic data for 5-fluoroindole-bound β-gly W33G. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgements

We thank Andrea Bazzoli, Yan Xia, Jittasak Khowsathit, Andrew Beaven, and Wonpil Im for valuable discussions and feedback. Support for the NMR instrumentation was provided by NSF Major Research Instrumentation Grant 9977422 and NIH Center Grant 5P20GM103418. Use of the IMCA-CAT beamline 17-ID at the Advanced Photon Source was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with the Hauptman-Woodward Medical Research Institute. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. This work was supported by grants from the National Institute of General Medical Sciences (8P20GM103420 and 5P30GM110761), the National Center for Research Resources (5P20RR017708), the National Science Foundation through XSEDE allocation ACI-1053575, the Human Frontier Science Program (J.K.), and the Alfred P. Sloan Fellowship (J.K.). S. Jimmy Budiardjo was supported by a Post Baccalaureate Research Education grant (2R25GM078441) and the Pharmaceutical Aspects of Biotechnology training grant (5T32GM008359) from the National Institutes of General Medical Sciences of the National Institutes of Health. Timothy J. Licknack was supported by the NSF REU Program (DBI-1156856). Michael B. Cory was supported by the Beckman Scholars Award from the the Arnold and Mabel Beckman Foundation.

Chapter 2

Druggable Pockets Exist in the Conformational Ensemble of Protein-Protein Interaction Sites Distinct from Inhibitor-Bound Crystal Structures

S. Jimmy Budiardjo[1], David K. Johnson[2], and John Karanicolas[3*]

[1] Center for Computational Biology, The University of Kansas, 2030 Becker Dr., Lawrence, KS 66047, [2]Computational Chemical Biology Core, The University of Kansas, 2034 Becker Dr., Lawrence, KS 66047, [3]Program in Molecular Therapeutics, Fox Chase Cancer Center 333 Cottman Ave, Philadelphia, PA 19111,

Introduction

Protein-protein interactions (PPIs) are involved in essentially every pathway in the human interactome. Therefore, modulation of PPIs offers enormous untapped potential for therapeutic intervention of diseases afflicting society. There have been great advancements in drug discovery in the last few decades with the advent of high throughput screening [71, 72], fragment-based drug discovery [73-75], and computational techniques [76] leading to potential drug candidates. Yet even with today's advancements, efforts in targeting protein-protein interactions are far from routine. This is made apparent by scarcity of drug candidates that have advanced into clinical trials over the last few decades. Protein-protein interactions have historically been deemed "undruggable". Just over a decade ago, James Wells deemed them the "high-hanging fruit" of drug discovery [77]. An archetype of protein-protein interactions is the B cell lymphoma 2 (Bcl-2) family of proteins that regulate apoptosis. Bcl-2 family members have arguably been the most heavily studied proteins involved in PPIs due to their role anti-apoptotic activity of cancer cells. However, it was not until 2015 which marked the first FDA approval of a selective small molecule inhibitor of Bcl-2 [78]. Developed by Abbvie and Genentech, Venetoclax received breakthrough therapy designation as a first-in-class treatment against chronic lymphocytic leukemia (CLL).

The difficulty in drugging proteins of this nature stems from their interaction surfaces, which are flat and relatively featureless [9]. In contrast, traditional drug targets such as enzymes or GPCRs contain well-defined deep binding pockets that make them amenable to structure-based techniques for inhibitor design. A key discovery revealed that small patches of residues, known as hotspots, contribute to the majority of the binding free energy between proteins making them amenable to small molecule inhibition without the need to occlude the entirety of

the interface [10, 11]. Comparison of the binding interfaces in crystal structures between the unbound state versus inhibitor bound or cognate protein bound state reveal that the relatively flat interaction surfaces undergo conformational rearrangements resulting in well-defined pockets, known as cryptic sites. Experimentally, NMR [79], multiple solvent crystal structures [80], and molecular tethering [81-83] have been used successfully to reveal these cryptic sites. However, exploring dynamics at the atomic detail necessary to capture new druggable pockets is limited with current experimental techniques.

For targets not amenable to these techniques, there has been an increased interest in recent years to use computational methods to reveal cryptic sites through implementation of many different flavors of molecular dynamics simulations. Analogous to experimental SAR by NMR, cosolvent simulations have been used to identify hotspot interfaces by including organic probe molecules (benzene, ethanol, propanol, imidazole, etc.) in solution that cluster on hydrophobic patches on the protein, and in some cases induce pocket formation [12, 84-87]. Long timescale MD simulations, on the order of hundreds of microseconds, have been used to build Markov state models that reveal cryptic binding sites confirmed by experiment [13, 88, 89]. Enhanced sampling techniques such as Accelerated Molecular Dynamics (aMD) apply a boost potential to the wells of the energy landscape reducing the barriers between states allowing efficient transitions potentially missed by conventional molecular dynamics [90]. Coupled to a pocket mining algorithm, aMD was used to determine druggability of each snapshot within a Bcl-xL ensemble. In this study, they found that 2% of the ensemble adopted large druggable pockets comparable to the properties Bcl-xL bound the to small molecule ABT-737. Mapping of cryptic site location alone is not sufficient in identifying druggable protein conformations so

using a combination of enhanced sampling and cosolvent methods has shown to improve identifying pockets with the magnitude of depth needed to bind a drug like compound [91, 92].

What the previous methods have in common is that they require intensive computational resources. We previously developed a non-MD based sampling approach implemented in the Rosetta software suite used to generate ensembles of low-energy conformations suitable for binding small molecules through biased simulations that is computationally less computationally demanding [14, 93, 94]. Rosetta uses a Monte Carlo search to explore backbone and sidechain degrees of freedom and we added a biasing potential to the standard Rosetta energy function to drive the ensemble towards pocket containing conformations, akin to enhanced sampling MD approaches. The biasing term can be interpreted as the energy associated with small molecule binding. Using this technique, we previously investigated cryptic pocket formation and found that the interfaces are more predisposed to forming pockets than the rest of the protein [14]. The Vajda group came to a very similar conclusion when analyzing NMR ensembles of PPI targets and found that predicted binding sites from the unbound ensemble overlap with the known binding sites from inhibitor or peptide bound structures [95]. In another study, we explored the determinants of selectivity within the Bcl-2 family members using our pocket-optimization sampling protocol. We found that cryptic pockets can be quite malleable as the same protein can adopt different conformations when bound to ligands of different chemical classes. Conversely, different proteins bound to the same ligand adopt similar pocket shapes [94]. With respect to their druggability, because inhibitor bound crystal structure only represent one low energy conformation in the ensemble, it is important to consider protein dynamics which, in principle, could reveal alternate low energy conformations that are suitable for the design of new drug candidates.

To explore this idea, we applied our Pocket-Optimization protocol to a set of four PPI targets. Bcl-xL, Mcl-1, MDM2, and IL-2 and compare conformational ensembles from simulation to known small molecule bound crystal structures. To do this, we next applied our newly developed high-throughput pocket shape comparison algorithm which uses a geometric analysis to map protein surface topography to compare cryptic site conformation similarity. Through this analysis we find that some proteins involved in protein-protein interactions can adopt druggable pockets that are distinct from currently known inhibitor bound complexes, which offers the potential for trapping distinct conformational states and has implications for selective targeting and the potential for novel small molecule drugs.

Results

Computational Workflow

We previously showed that the Pocket-Optimization protocol generates low-energy ensembles that overlap with ensembles generated without the biasing term using the Rosetta "relax" protocol [14, 94]. The Rosetta software suite's energy function is a combination of physics-based force field terms and knowledge based statistical potentials [96]. A systematic comparison between a purely Molecular Mechanics based force field Amber and Rosetta perform similarly well in protein structure evaluation [97]. As such, as a basis for determining physiological relevance of pocket optimized conformations, we also generated ensembles using conventional unbiased MD simulations as it provides a more thermodynamically rigorous ensemble for comparison.

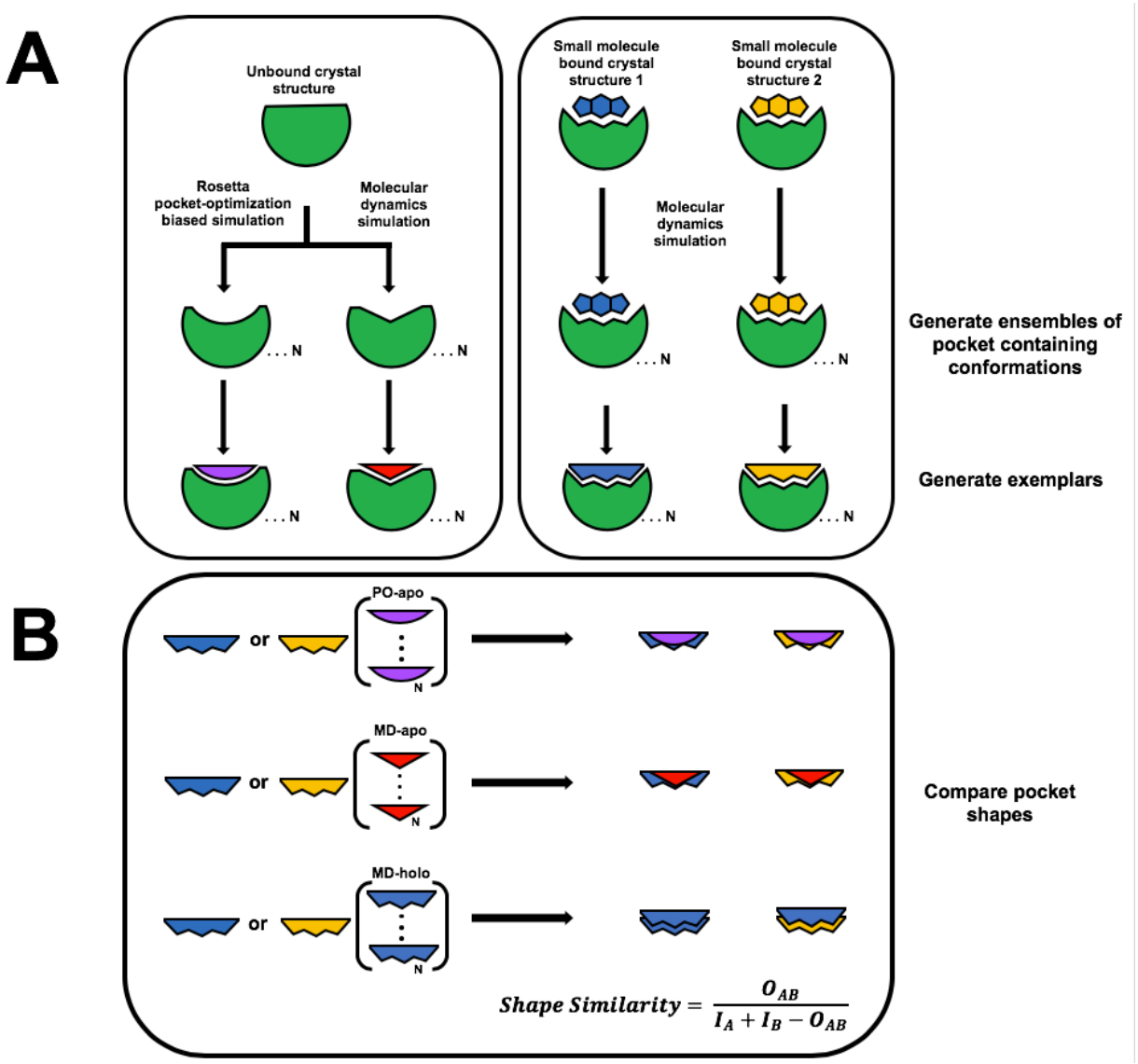


Figure 2.1. Computational Workflow: (A) Ensemble generation from the unbound protein using Pocket-optimization and molecular dynamics to find pocket containing conformations (*left*). Ensemble generation of the inhibitor bound proteins by molecular dynamics to generate pocking containing conformations (*right*). Exemplars are generated for each snapshot as a pocket shape descriptor (*purple, red, blue, yellow*). **(B)** Pocket shapes are quantitatively compared by the volume overlap of each exemplar using ROCS which calculates the ShapeTanimoto where O_{AB} is the volume overlap of exemplars A and B, I_A is the self-volume of exemplar A, and I_B is the self-volume of exemplar B.

For each system we set up 4 simulations: the unbound protein through Pocket-Optimization, unbound protein through molecular dynamics (**Figure 2.1A** left), and two inhibitor bound systems through molecular dynamics (**Figure 2.1A** right). Exemplars were generated for each

snapshot of the ensemble as a geometric pocket shape descriptor as a basis for conformational comparison (**Figure 2.1A**). We previously introduced the concept of an "exemplar" which is a geometric descriptor of a pocket's volume and shape [93, 94]. Exemplars are built by placing the protein onto a grid and filling the void volume of a pocket with a collection of spheres. The spheres are not constrained to physical bond lengths or particular atomic radii and can be seen as an "idealized" ligand that represents the perfect complementary surface structure. By representing the pocket in a ligand-centric manner, we can compare shapes in a high-throughput quantitative matter using available virtual-screening shape comparison tools.

Rapid Overlay of Chemical Structures (ROCS) calculates the three-dimensional similarity as the volume overlap between chemical entities [98]. The "chemical entities" in our case being exemplars of pockets of pocket shapes. Distances are reported by the ShapeTanimoto as calculated with the following equation:

$$shapeTanimoto = \frac{O_{AB}}{I_A + I_B - O_{AB}} \qquad \text{(Equation 2.1)}$$

Where O_{AB} is the volume overlap of exemplars A and B, I_A is the self-volume of exemplar A, and I_B is the self-volume of exemplar B. Exemplars from each ensemble are compared to the experimentally determined crystal structure to determine (**Figure 2.1B**). Likewise, exemplars between all snapshots of each ensemble (all vs all) can also be compared to determine the pocket space sampled.

Systems

Protein-protein interactions can be classified based on the secondary structure of interacting partners. It has been suggested that PPIs involving a globular protein and a single linear or helical peptide element of the other protein are more druggable [99-101]. The binding peptide typically engages with two or three hotspots or sub-pockets on the protein. Inhibitors typically mimic the interactions of the peptide sidechains in the pocket. We selected three systems from this category of PPI: Bcl-xL, Mcl-1 and MDM2 (**Figure 2.2A-C**). On the other hand, PPIs involving two globular proteins have proven more difficult to target. The interfaces are discontinuous and involve tertiary structural elements on both proteins. Assigning which protein to assign as the "receptor" vs "ligand" to target is arbitrary and both proteins could be probed for cryptic pockets. Under this category we also simulate IL-2 (**Figure 2.2D**).

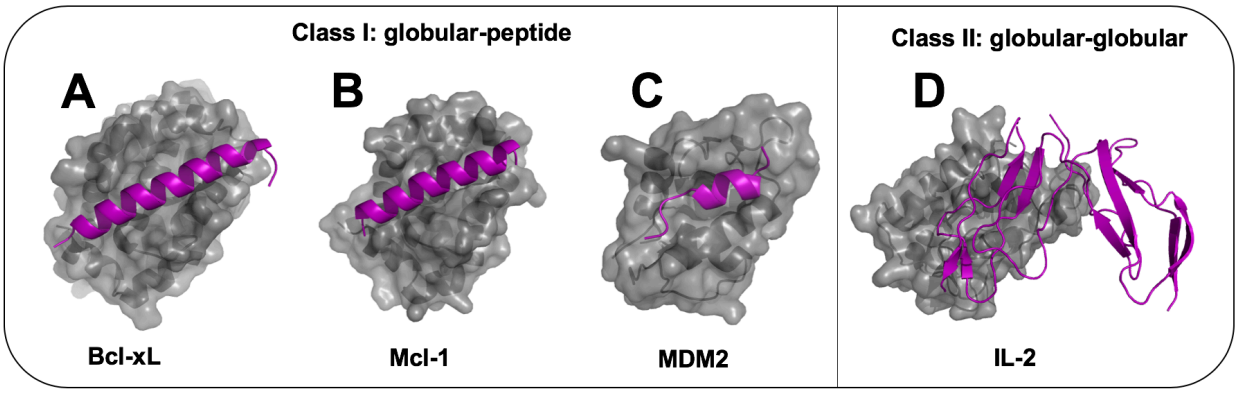


Figure 2.2. Protein-Protein Interaction Classes: Protein targets (*grey surface*) with cognate protein partner (*magenta cartoon*) for two different classes of PPIs. Class I - Globular protein with a single peptide **(A)** Bcl-xL bound to BH3 peptide BAD PDBID:2BZW **(B)** Mcl-1 bound to BH3 peptide Bim PDBID:2PQK **(C)** MDM2 bound to p53 PDBID:1YCR. Class II – globular protein with another globular protein **(D)** IL-2 bound to IL-2Rα PDBID:1Z92.

B-cell lymphoma extra-large (Bcl-xL) is a member of the Bcl-2 family of proteins which regulate apoptosis. The delicate balance between pro and anti-apoptotic activities determines a

cell's fate. When Bcl-xL is upregulated, antiapoptotic activity contributes a cancer cell's ability to circumvent cell death. In the bound state, Bcl-xL forms a long cleft containing two hydrophobic pockets that interacts with a single helical peptide (BH3 domain) of proapoptotic partners. BH3 peptides burry four hydrophobic residues into pockets designated P1-P4 [102]. Small molecule inhibitors of Bcl-xL are large (~800 Da) elongated molecules that span the hydrophobic groove and engage hotspots P2 and P4 (**Figure 2.3A**) [103]. This is reflected in the shape of the exemplar which encompass the small molecule and fill the volume of the pocket.

Induced myeloid leukemia cell differentiation protein (Mcl-1) is another Bcl-2 family member with antiapoptotic activity. Upregulation of Mcl-1 contributes to Bcl-xL inhibitor resistance [42]. Although sequence identity between Mcl-1 and Bcl-xL is only ~18%, they share the same structural fold and binding groove but have different selectivity profiles for BH3 peptides. Relative to Bcl-xL, the P2 pocket is highly dynamic while the P4 pocket is more rigid [102]. As such, inhibitors identified for Mcl-1 only engage P2 but penetrate deeper (**Figure 2.3B**) [104]. The pocket shape is represented by a compact clustering exemplar.

Mouse double minute 2 homolog (**MDM2**) is an E3 ubiquitin-protein ligase that down regulates tumor suppressor p53 through ubiquitin mediated protein degradation and thus is a target in cancer therapy. Like the Bcl-2 family proteins, the PPI interface involves a globular protein (MDM2) that binds to a single short peptide p53 **(Figure 2.3C**). The MDM2 pocket forms a wide bowl-like conformation in which the peptide of p53 which presents three hydrophobic residues close in primary space [105]. Small molecule inhibitors of MDM2 form trident that burry into the same three hydrophobic hotspots [106].

Interleukin-2 (IL-2) is an immune response regulator involved in T-cell activation and proliferation upon antigen presentation. Suppression of the immune response during allograft

transplantation by IL-2 inhibition is of interest for reducing tissue rejection [107]. IL-2 forms a ternary complex with membrane bound receptors IL-2Rα, IL-2Rβ, and IL-2Rγ. Of therapeutic interest is the disruption of the interface between IL-2/IL-2Rα. Unlike the aforementioned PPIs that involve a globular protein with a single peptide, the interaction surface between IL-2 and IL-2Rα involves tertiary structure elements between two globular proteins (**Figure 2.2D**). As such, interface topography is more complex and discontinuous between hotspots leading this class of PPI to be less druggable. The binding mode of known inhibitors span two regions on the surface (**Figure 2.3D**). The first is a pre-formed cluster of residues that form polar interactions and a salt bridge with E62 and the second, a dynamic hydrophobic pocket only present in the small molecule bound structure as a result of a groove forming through rotation of F42 [107]. Because E62 is already posed for binding, we focus our analysis on the dynamic hydrophobic pocket that forms upon complexation.

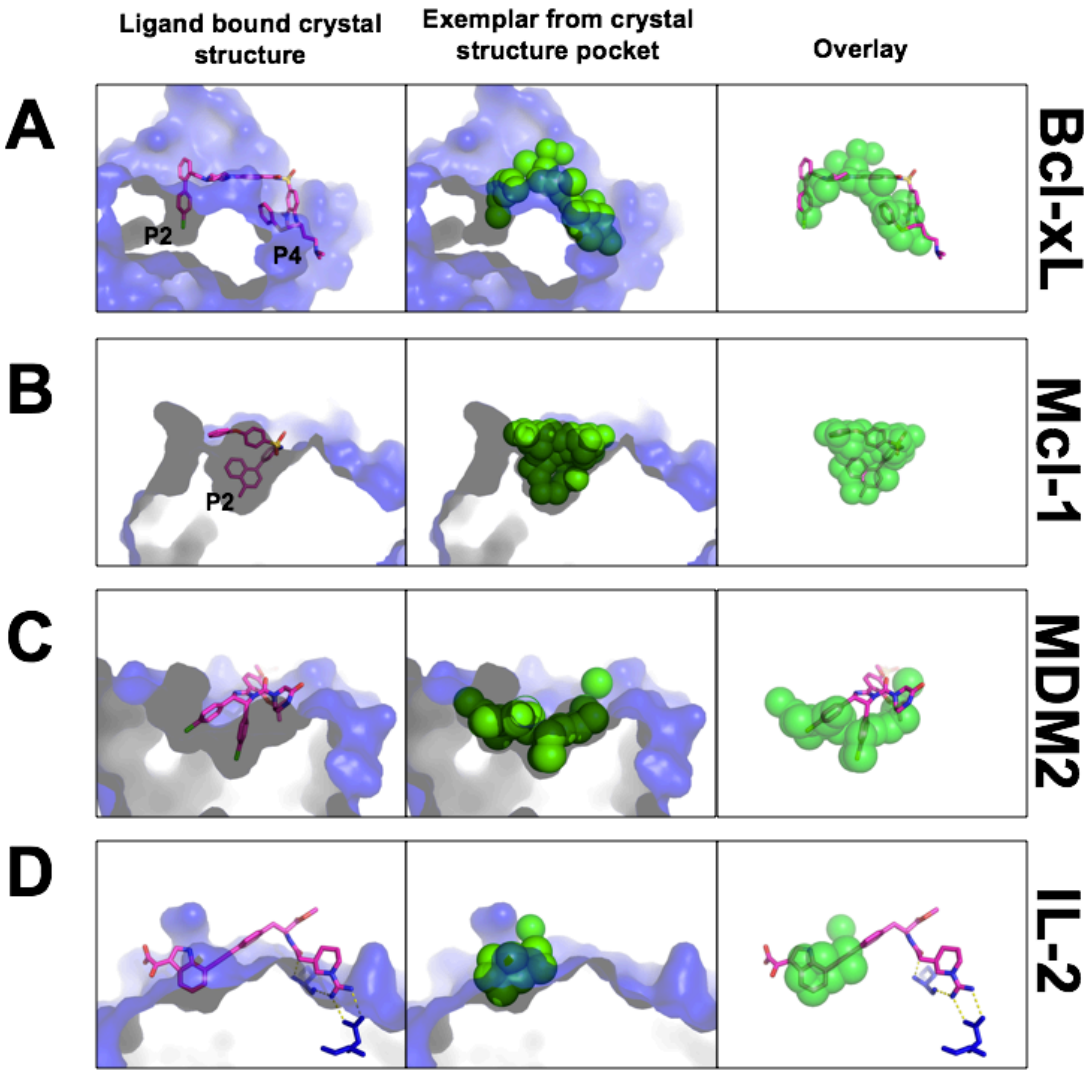


Figure 2.3. Representative small-molecule bound crystal structures and exemplars generated: Crystal structure conformations (*blue, surface*) are shown in the bound state with inhibitors (*magenta sticks, left*), exemplars mapping the pocket shape (*green spheres, middle*) and an overlay of the inhibitor and exemplar (*right*) for Bcl-xL 2YXJ **(A)**, Mcl-1 4OQ5 **(B)**, MDM2 4HG7 **(C)**, and IL-2 1M49 **(D)**. For IL-2 the interface consists of two regions, a transient hydrophobic pocket and a preformed cluster of residues that form salt bridge interactions (*yellow dotted lines*).

Pocket Dynamics

One of the most important factors in determining pocket druggability is its volume. Pockets with volumes greater than 200 $Å^3$ have previously been determined to be the size necessary to bind a small molecule [90, 108, 109]. As the most reliable metric for druggability, we first determined the pocket volumes of the orthosteric site of each snapshot of the simulations using our "deep-volume" algorithm. **Figure 2.4** shows distributions of pocket volumes from each ensemble, Apo-Pocket-Optimized (red), Apo-MD (blue), Holo-1-MD (solid line), Holo-2-MD (dashed line). Pocket volumes of the static crystal structure representing the starting structures for MD-holo-1 and MD-holo-2 are shown in the vertical solid and dashed lines, respectively. Unsurprisingly, the peak of the distributions for the bound simulations are above 200 $Å^3$ and also sample larger volumes than the apo simulations. The exception is IL-2 since the transient pocket that forms only needs to encompasses half of the small molecule inhibitor and the pocket in the bound structures start just above 100 $Å^3$. Starting from the apo structure, the orthosteric site interface is able to visit conformations that exceed the 200 $Å^3$. Remarkably, in all cases the distributions of volume between the unbound ensembles (PO or MD) overlap with ensembles where the small molecule inhibitor is present during the simulation. This also not due to pocket collapse from complete dissociation of the inhibitor since they remained bound during the entirety of the simulations. However, the degree of overlap depends on the protein system. For Bcl-xL the most of the conformations contain no detectable pocket but a small subset of the ensemble opens as large as the inhibitor bound state. Although the distributions for Mcl-1 show more pocket containing conformations for Mcl-1 greater than 200 $Å^3$, the overlap with the bound distributions is less than Bcl-xL. This could be due to the long shallow nature of the Bcl-xL interface versus the need for much deeper pocket formation in Mcl-1. Both IL-2 an MDM2 both

sample a large span of volumes similar to that of the bound ensembles. It is notable that the distributions between Pocket-Optimization and molecular dynamics are quite similar indicating that they are sampling similar landscapes.

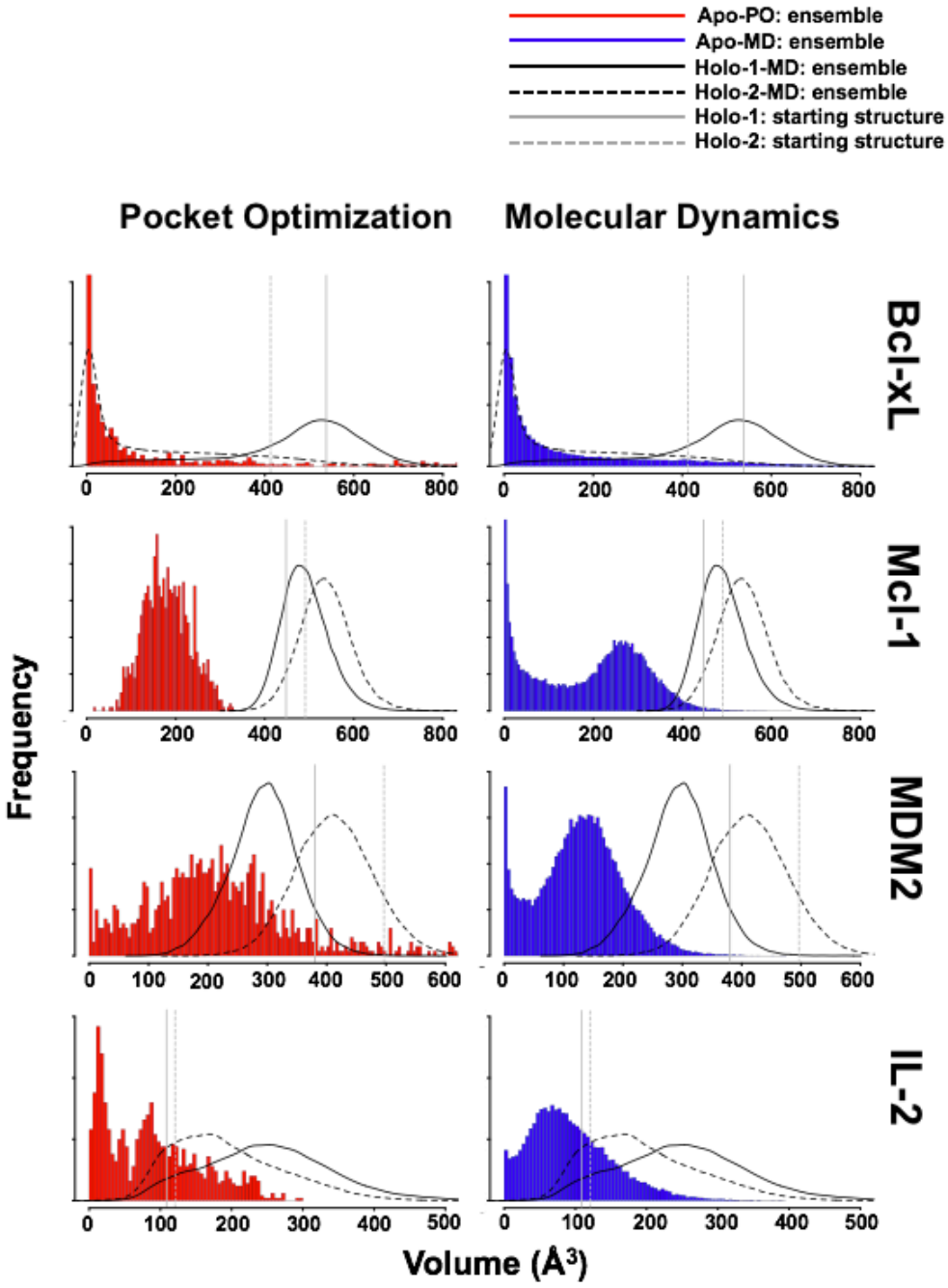


Figure 2.4 Pocket volumes from each simulation: Pocket volumes for each simulation for Apo-Pocket-Optimization (*red*), Apo-MD (*blue*), Holo-1-MD (*solid line*), Holo-2-MD (*dashed line*). Vertical lines represent the volumes of the starting structures for the inhibitor bound simulations. Bcl-xL: 2YXJ solid line, 3ZLN dotted line. Mcl-1: 4HW3 solid line, 4OQ5 dotted line. MDM2: 4HG7 solid line, 4MDN dotted line. IL-2: 1M49 solid line, 1PW6 dotted line. Unbound ensembles show overlap of pocket volumes sampled in the inhibitor bound ensembles.

Exemplar distance distributions

Since the ensembles in all cases are able to visit states that are of similar size to the inhibitor bound ensembles, this poses to another question. For the pockets that do form, how similar are they to the experimentally determined inhibitor bound crystal structure? We generated exemplars for each snapshot and used ROCS to determine their similarities. Pocket similarities are expressed as the shapeTanimoto which describes volume overlap between exemplars. Exemplar score ranges between 0-1, with 1 being an identical match. Distributions of pocket distances to inhibitor bound crystal structures are shown in (**Figure 2.5)**. In the holo simulations (black lines), the peak of the distance distributions range between 0.4-0.6 indicating that a majority of the time the protein samples conformations slightly altered from the inhibitor bound states. The long tails on both sides of the distribution suggest that the interfaces are quite dynamic and there is flexibility to deviate from the crystal structure. For IL-2 the bound simulations have a wide distribution which is indicative of the pocket forming similar shapes to the unbound simulation much more readily than the other proteins. Remarkably, the apo simulation distances overlap with the inhibitor bound ensembles. This suggests that both ensembles are visiting overlapping conformational space without any direct knowledge of a particular ligand from the unbound state.

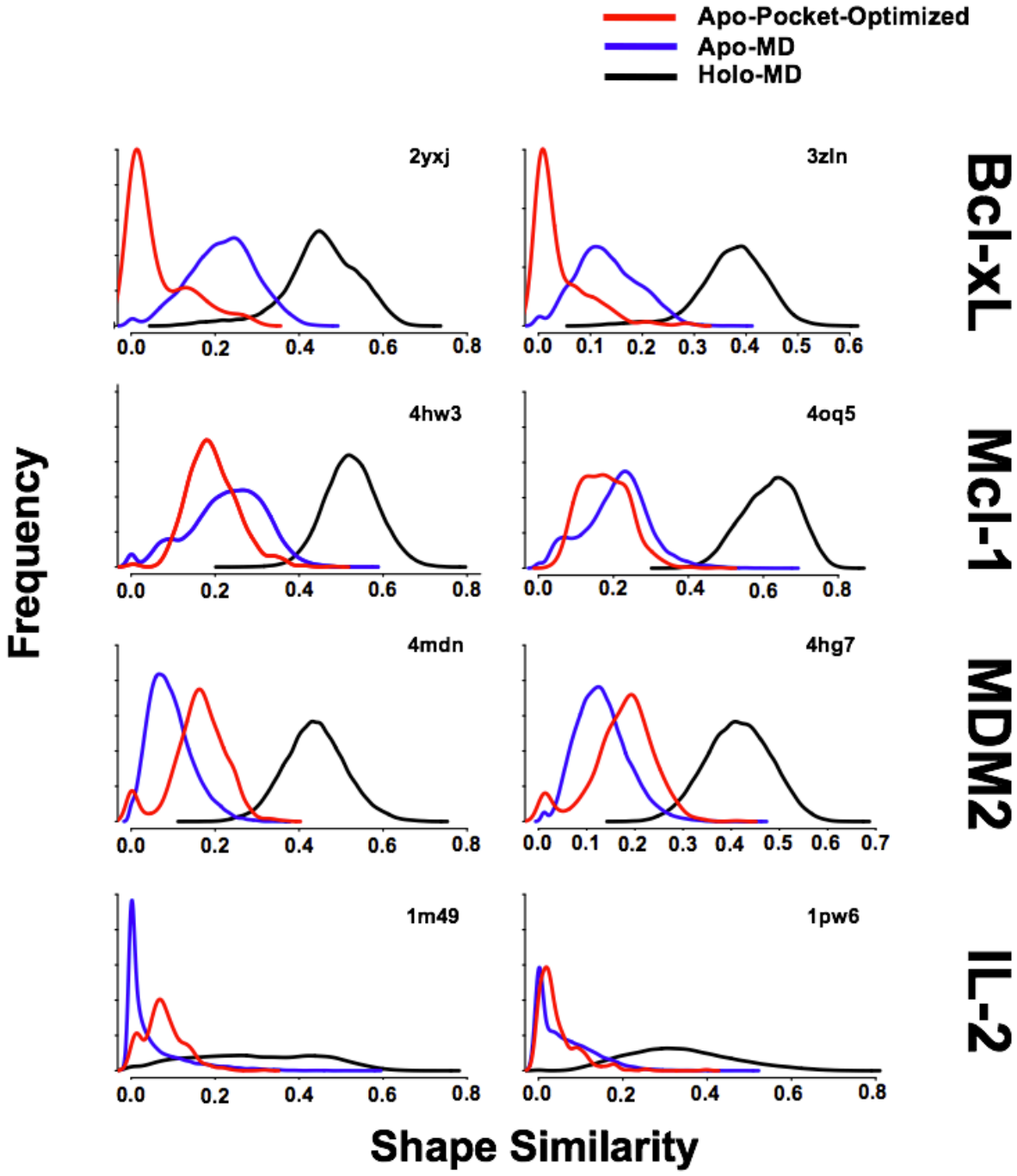


Figure 2.5 Exemplar distance distributions: Exemplar distances calculated with respect to the inhibitor bound crystal structures for Apo-Pocket-Optimized (*red*), Apo-MD (*blue*), Holo-MD (*solid line*). Unbound ensembles show overlap of pocket similarity sampled in the inhibitor bound ensembles.

From the Apo-Pocket-Optimization ensembles we pulled the protein snapshot with the highest scoring exemplar (**Figure 2.6**). To compare the conformational transitions associated with binding we superposed the unbound starting structure (magenta) with the protein conformation of the top scoring exemplar from the Pocket-Optimized simulation (green) to the inhibitor bound conformation (cyan). Features of the bound protein which did not exist in the starting unbound structure are recapitulated. Many of the conformational changes involve sidechain rearrangements that open the pocket. For Bcl-xL rotation of F97 opens part of the pocket which would otherwise clash with the ligand (**Figure 2.6A**), additionally helix winding which rotates W137. As Mcl-2 inhibitors engage a different pocket more deeply than Bcl-xL inhibitors, requiring more backbone rearrangement as seen in a helix shift which can be captured by M231 (**Figure 2.6B**). In MDM2 we see a concerted outward rotation of M62, V93, H96, Y100 residues to open the pocket (**Figure 2.6C**). For the transient pocket of IL-2, we see the same rotamers necessary to open the groove upon rotation of F42 and M39 (**Figure 2.6D**). These features show that the unbound ensemble can access the bound states. This is in line with our previous study which suggested that pocket shape and thus selectivity are encoded in surface fluctuations akin to a conformational selection model of binding.

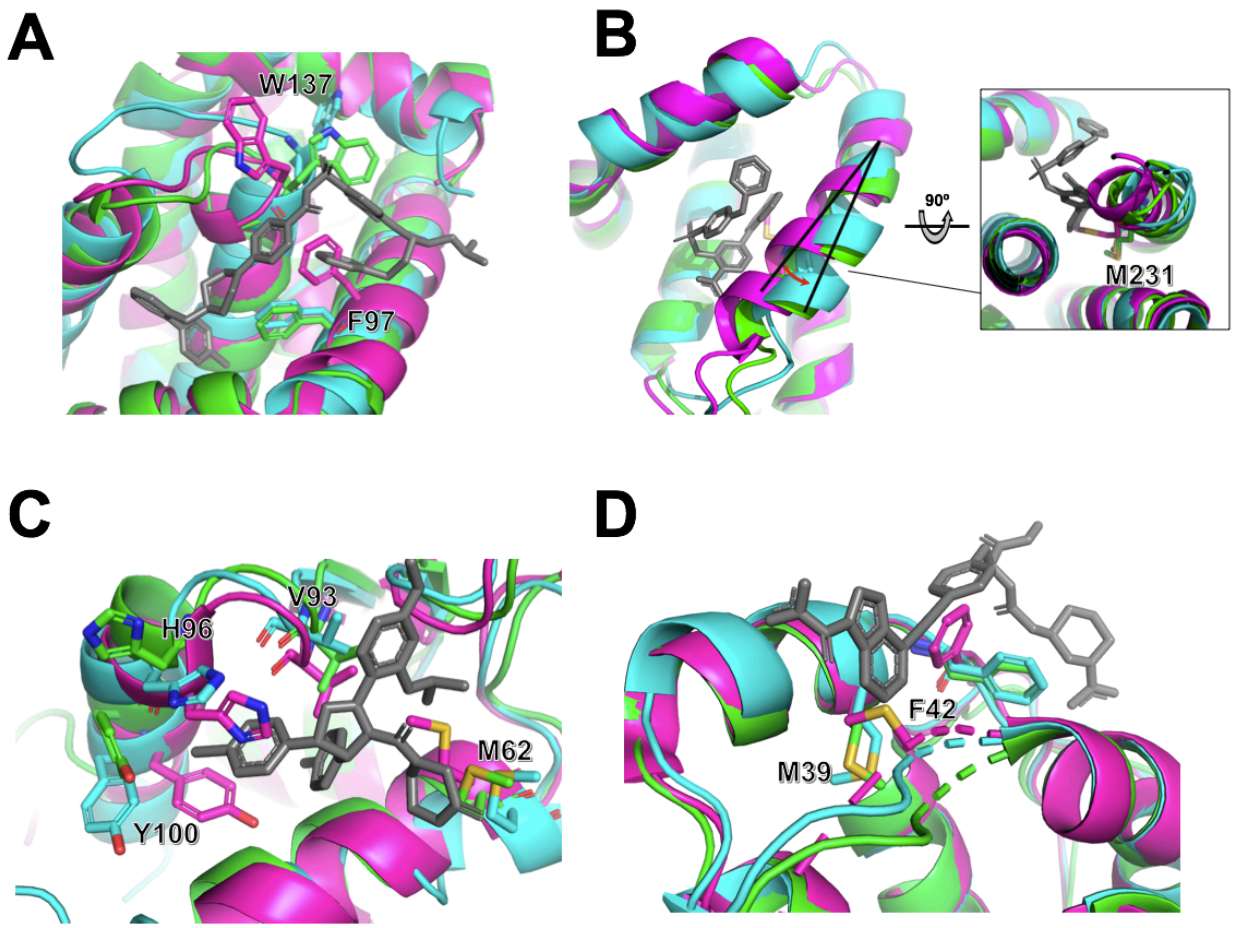


Figure 2.6. Bound conformations are recapitulated in the unbound ensemble: Conformations of the top scoring exemplar from Apo-Pocket-Optimization ensembles (*green*) are superposed with the inhibitor bound conformation (*cyan*) and the unbound starting structure (*magenta*) for Bcl-xL 2YXJ **(A)**, Mcl-1 4OQ5 **(B)**, MDM2 4HG7 **(C)**, and IL-2 1M49 **(D)**.

MDS analysis

We next compared conformational "pocket space" to explore the differences between ensembles of each simulation. The distances between exemplars from each simulation were calculated and used for multidimensional scaling (MDS) analysis to produce a two-dimensional visual projection that preserves the distance between every pair of exemplars (**Figure 2.7**). For each simulation, 100 representative snapshots evenly spaced through the trajectory were selected

for MDS analysis. Since not all conformations contain pockets and thus exemplars, each plot contains less than 100 points for each simulation. For each protein system, exemplars were also generated for all known experimentally determined inhibitor bound crystal structures and included in the MDS analysis (points 1-9, A-Z). In all cases, ensembles from the apo simulations (PO red, MD blue) show extensive overlap indicating that the conformations generated from Pocket-Optimization are indeed physiologically relevant. For the Bcl-2 family members (class I PPI) the unbound simulations do show some overlap with known inhibitor bound crystal structures indicating that the ensembles are able to visit inhibitor-bound like states. There is much less agreement to the ensembles from the inhibitor bound simulations as they sample conformations further away from the static inhibitor bound structures. Of particular interest, there are regions in the unbound simulations forming that sample distinct pocket shapes. Both pocket-optimization and MD are finding the same low-free energy pocket-containing conformations that have yet to be captured by any known inhibitor. This observation is notable because these conformations could offer underexplored regions of conformational space that have not been targeted by small molecule inhibitors. For MDM2, another class I PPI, shows similar behavior to the Bcl-2 proteins. Interestingly for IL-2, the sole class 2 PPI from this study, the unbound simulations show complete coverage of both the inhibitor bound crystal structures and the inhibitor bound ensembles indicating that the pocket that forms is always of the same shape.

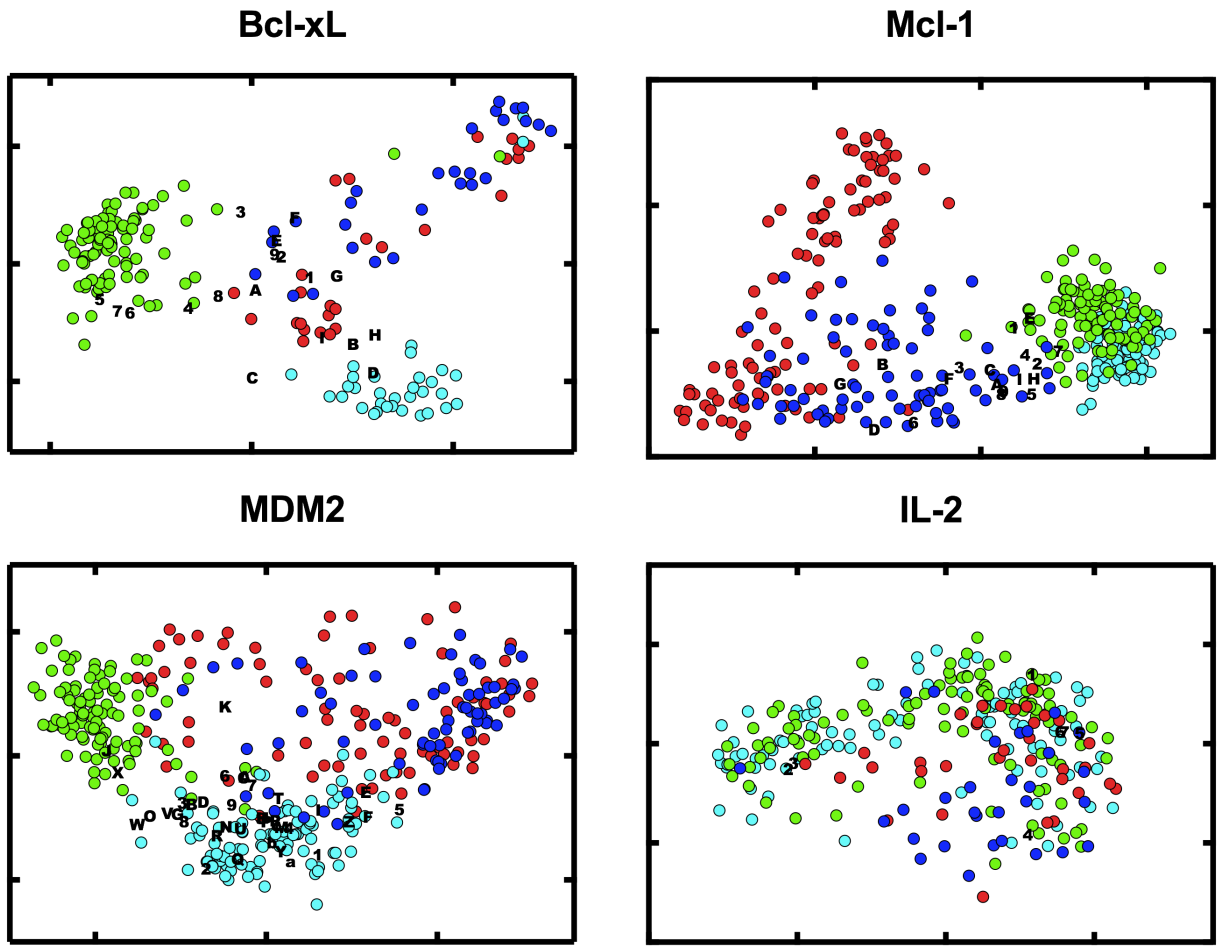
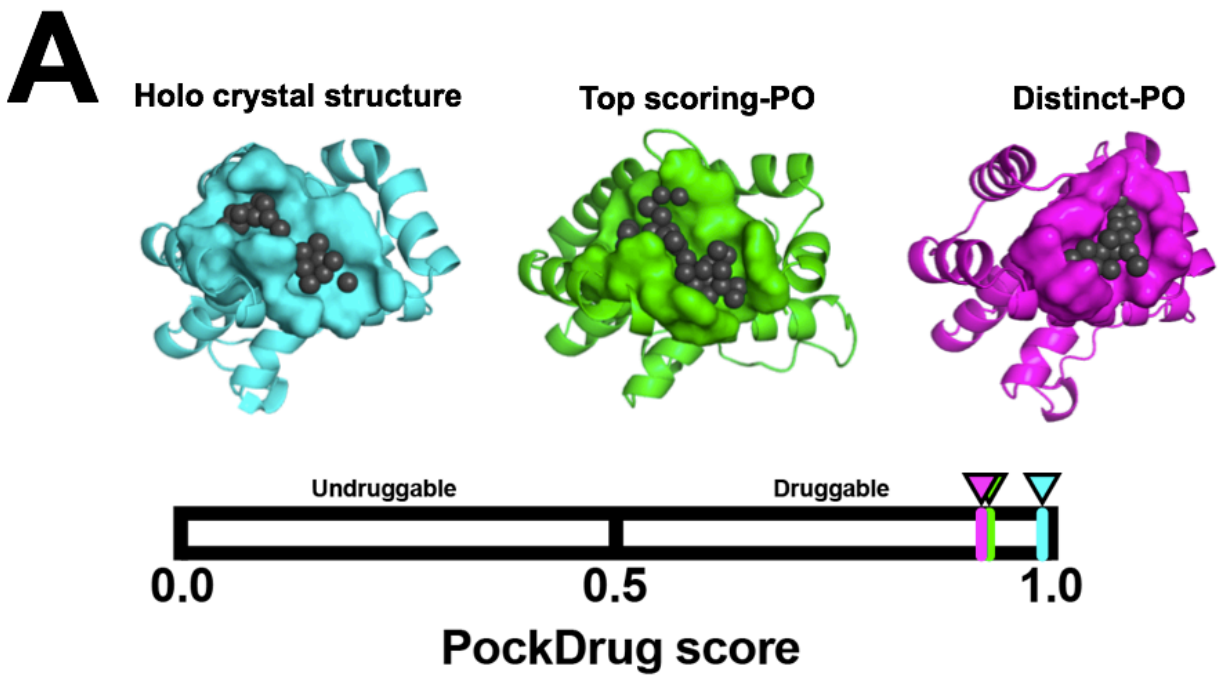


Figure 2.7. Conformational Pocket Space: MDS analysis projecting pocket space onto a 2D map where the distances between exemplars are preserved. Apo-Pocket-Optimization (*red*), Apo-MD (*blue*), Holo-1-MD (*cyan*), Holo-2-MD (*green*). Also included are exemplars from other known inhibitor bound crystal structures (*1-9-A-Z*).

Distinct Pocket in Bcl-xL

We next sought to identify a pocket containing conformation from the distinct region on the MDS map and determine its druggability relative to the known inhibitor bound conformation. Instead of visually choosing a conformation from the MDS map, we instead used the exemplar distances to compute the distinct conformation by taking the ratio of the distance to the closest

crystal structure to the distance of the closest MD structure then ranking them. This gives us a conformation that is far from known inhibitor bound crystal structures and close to the conformations sampled by MD. We applied this to the unbound Bcl-xL ensemble from Pocket-Optimization and the conformation is shown in **Figure 2.8A** (magenta).



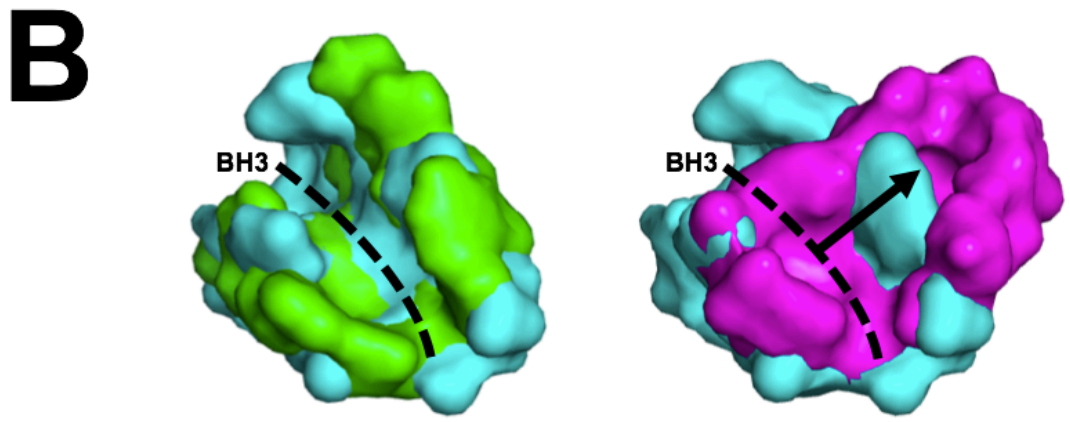


Figure 2.8. The unbound ensemble of Bcl-xL visits a highly druggable distinct state: (A) Conformations of the inhibitor bound structure (*cyan*), the top scoring most similar snapshot from Apo-Pocket-Optimization (*green*), and distinct state from Apo-Pocket-Optimization (*magenta*). Each conformation was submitted to the PockDrug server to assess druggability. Pocket shell residues identified through PockDrug analysis are shown in surface with exemplars in black spheres. PockDrug scores >0.5 are deemed druggable. Each conformation scores high in druggability with scores of 0.92 for the distinct state, 0.93 top scoring state, 0.98 inhibitor bound state. **(B)** Pocket shell residues superposed for the top scoring and inhibitor bound conformations (*left*) and the distinct state and inhibitor bound conformation (*right*). The canonical BH3 binding groove is highlighted as a dotted line. The distinct state opens a pocket orthogonal to the direction of the BH3 binding groove.

As a reference, the inhibitor bound conformation and top-scoring conformation to the bound are shown in cyan and green, respectively. Druggability of each pocket was determined using PockDrug which determines a druggability score between 0.0 and 1.0 with a pocket >0.5 deemed as druggable [110, 111]. The distinct pocket containing conformation adopts a highly druggable state with a druggability score (0.92) similar to that of the top scoring conformation (0.93) and inhibitor bound crystal structure (0.98). Pocket shell residues from PockDrug analysis were extracted and superposed. The top scoring conformation adopts the canonical BH3 binding groove that would normally become occluded by the small molecule inhibitor (**Figure 2.8B** left). On the other hand, the distinct conformation contains two characteristics that make it appealing as a target. First, the pocket forms in a manner that precludes BH3 binding. Second, it overlaps with the P4 pocket yet extends in the direction orthogonal to the canonical BH3 peptide site making it a new state that has yet to be drugged (**Figure 2.8B** right, arrow). An inhibitor that binds to the distinct pocket would not be based entirely on mimicry of the BH3 peptide moieties since it would engage an entirely new region of the protein.

Discussion

Discovering novel small molecules that inhibit protein-protein interactions remains an unsolved problem. The dynamic nature of their binding sites presents an enormous hurdle for structure-based drug design as the features that define a pocket are not present in the static unbound crystal structures. Moreover, for the proteins where the experimentally determined structures are known, the pockets of small molecule inhibitor bound proteins often times do not resemble the pocket when bound to its cognate protein partner. Although daunting, it also suggests that the interactions sites are malleable offering the opportunity to target undrugged states. Additionally, the use of previously identified inhibitors or pharmacophores based on side chains of cognate partners as scaffolds narrows the scope of potentially new molecular entities. It is therefore imperative to consider dynamics of the interface in order to broaden the success rate of discovering new drug candidates by finding new alternate highly druggable states.

In this study, we sought to explore pocket dynamics of PPIs and the "pocket space" available at each interface. Using our Pocket-Optimization protocol we were able to rapidly generate ensembles low energy of protein conformations containing pockets. We then applied our high throughput geometric pocket shape comparison protocol and found two main conclusions. First, pocket shapes derived from Pocket-optimization ensembles overlap with pocket shapes generated from ensembles including the inhibitor. This suggests a conformational ensemble like model of binding since the protein can adopt a particular shape without any knowledge of one ligand. When comparing the top scoring conformation to the inhibitor bound crystal structure, the features necessary for opening a pocket emerge that resemble the bound state. Secondly, when considering pocket space more globally, there are regions distinct from any known available inhibitor visited by the pockets. We argue that these states will give us the

opportunity to attack underexplored new protein conformational states and inhibitor chemical space as seen with a new distinct highly druggable conformation of Bcl-xL. Additionally, this could reduce off-target effects. For proteins that share overlapping selectivity profiles like the Bcl-2 family members, we can exclude the overlapping pocket space and only choose distinct conformations.

The caveat encountered is not all PPIs are created equal. We previously introduced two classes of PPIs. Class I being a globular protein-helical peptide interaction such as Bcl-xL, Mcl-1, and MDM3. Our sole representative of a globular protein-globular protein interaction with IL-2 showed different characteristics to the class I systems. There was much more overlap between unbound and bound ensembles. In addition, the MDS analysis shows that there was complete overlap between all ensembles and the inhibitor bound states indicating that when the pocket opens, it is the same pocket. The lack in alternate states could also be reflected in the difficulty in targeting class II vs class I PPIs, the interfaces of which are patchier and more discontinuous. Class II PPIs may need alternative strategies that rely on more traditional ligand centric methods for optimization like new *in silico* combinatorial chemistry libraries with drug-like properties.

Looking forward, Pocket-Optimization adds to the available computational repertoire to explore pocket opening dynamics with the advantage of having a low computational cost relative to other methods. In the current study we limited our analysis with a priori knowledge of the orthosteric site of previously discovered inhibitors. In principle however, pocket-optimization could also be readily coupled to any of the pocket identification protocols that are currently available such as the co-solvent techniques. Additionally, this technique could also be applied to novel allosteric sites to identify new ligands such as stabilizers or activations, in addition to inhibitors.

The real test of course is to find inhibitors for these distinct states. Our high throughput geometric pocket comparison protocol using exemplars can also be used in turn as a platform for virtual screening. Exemplars in the context of this study were used as a pocket shape descriptor but they can also be seen as the ideal ligand shape that would fit a particular pocket. One would first identify distinct conformations through finding unique exemplars within the ensemble. Then using this exemplar, one could use it as a template for ligand based virtual screening by not only including the shape and volume overlap of spheres, but also including chemical features such as hydrogen bond donors and acceptors. With the advances in computational power and the boom in machine and deep learning techniques applied to modulating protein-protein interactions [112], we believe new avenues in computational drug discovery will be available to address these challenging targets.

Methods

Pocket-Optimization simulations:

Starting from the unbound structure of each protein, an ensemble of pocket containing conformations was generated using a modified version of the "relax" protocol in the Rosetta software suite [96]. Input structures of the apo proteins were taken from the Protein Data Bank: Bcl-xL (PDB 2M03), Mcl-1 (PDB 2MHS), MDM2 (PDB 1Z1M), IL-2 (PDB 1M47). Missing densities were built using homology modeling. The relax protocol uses a Monte Carlo search to sample backbone and side chain degrees of freedom. We have previously described a modified version that enriches sampling to include low energy pocket-containing conformations [14, 94]. Briefly, a biasing energy term consisting of the "deep pocket" volume multiplied by a

proportionality constant was added to the standard Rosetta energy function which enhances conformations that contain surface pockets. From each unbound protein conformation of Bcl-xL, Mcl-1, MDM2, and IL-2, 1,000 independent trajectories were run. For this study, target residues for pocket optimization were selected manually by choosing residues making contacts with inhibitors in the small molecule-bound crystal structures.

Molecular Dynamics Simulations:

Input structures were taken from the Protein Data Bank: apo Bcl-xL (PDB 2M03), small molecule bound Bcl-xL (PDB 2YXJ and 3ZLN), apo Mcl-1 (PDB 2MHS), small molecule bound Mcl-1 (PDB 4HW3 and 4OQ5), apo MDM2 (PDB 1Z1M), small molecule bound MDM2 (PDB 4MDN and 4HG7), apo IL-2 (PDB 1M47), small molecule bound MDM2 (PDB 1M49 and 1PW6). Missing densities were built using homology modeling. Input structures for molecular dynamics (MD) simulations were prepared using CHARMM-GUI Input Generator (http://www.charmm-gui.org) [113]. Each PDB was immersed in a cubic water box of 65 Å x 65 Å x 65 Å. KCl concentration set to 0.15 M. CHARMM-GUI uses CGenFF to generate parameters for the small molecule inhibitors.

Simulations were run using NAMD2.10 [114] simulation package in NPT (constant particle number, pressure, and temperature) ensembles at 310 K and 1 atm using CHARMM all-atom force field [115] parameter set CHARMM36, with a dihedral cross-term CMAP and a TIP3 water model [116], van der Waals interactions smoothly switched off 10-12 Å by a force switching function [117], electrostatic interactions calculated using particle-mesh Ewald method and a 2-fs timestep [118]. Temperature and pressure were kept constant by Langevin dynamics and Nose-Hoover Langevin piston methods, respectively. The systems were equilibrated for 5-ns

with a decreasing RMSD harmonic-restraint potential on all heavy atoms by decreasing the force constant (10, 5, 2.5, 1, 0.5 kcal mol^{-1} $Å^{-2}$). Production runs were conducted for 100 ns of simulation time.

Pocket Volume Calculation

"Deep" pocket volumes were calculated within the Rosetta software suite as previously described [14]. Based on the same target residues used for the pocket opening simulations, a grid is placed around the target residue. Grid points are marked as solvent, surface, or protein and a linear search is conducted to find "Surface-Solvent-Surface" events. Pocket grid points that are within 2.5 Å of solvent are not included in the calculation and the remaining contiguous points are clustered and defined as "deep pocket".

Geometric Description of Pocket Shapes: Exemplars

Backbones of each snapshot from every simulation were aligned to the unbound starting structure to orient the orthosteric site in the same position. Exemplars were then built at each orthosteric site of each snapshot as previously described [93, 94]. Briefly, Exemplars are built from the "deep volume" described above by filling the unoccupied space with carbon atoms using a greedy algorithm. Atoms are not constrained based on real physical chemical bond lengths or connectivity but the center of atoms are placed no less than 1.7Å apart. Atoms are then clustered within 5Å such that if a target residue is flanked by two pockets it will be included as one exemplar.

Comparing pocket shapes: ROCS scoring

Rapid Overly of Chemical Structures (ROCS) software (OpenEye Scientific Software, Santa Fe, NM) is a virtual screening tool used to optimally superpose and calculate the volume and chemical overlap between two molecules[98]. ROCS uses the sum of atom centered Gaussians to represent molecular shape and volume. Additionally, the color descriptor represents chemical groups and uses another set of Gaussians. Since ROCS is primarily designed for the purposes of virtual screening, we turned off two functionalities for analysis and simply scored the volume overlap between exemplars. For exemplar shape comparisons, we turned off the superposition optimization and used the "score only" functionality since exemplars were already spatially oriented for comparison by generating exemplars from backbone aligned snapshots. Additionally, the color score was turned off since we are only reporting on pocket shape.

Druggability Prediction

Druggability was determined using the PockDrug web server (http://pockdrug.rpbs.univ-paris-diderot.fr/) [110, 111] by uploading each protein with its associated exemplar and using the proximity method with a threshold of 5.5Å to the exemplar.

Acknowledgements

We thank Andrew Beaven and Andrea Bazzoli for methods guidance and useful discussions. We are grateful to OpenEye Scientific Software (Santa Fe, NM) for providing an academic license for use of ROCS.

Chapter 3

Colicin E1 Fragments Potentiate Antibiotics Through Occlusion of TolC

S. Jimmy Budiardjo[1], Jacqueline J. Deay[2], Anna Calkins[3], Virangika K. Wimalasena[2], Julie S. Biteen[3] and Joanna S.G. Slusky[1,2*]

[1]Center for Computational Biology, The University of Kansas, 2030 Becker Dr., Lawrence, KS 66045-7534 [2]Department of Molecular Biosciences, The University of Kansas, 1200 Sunnyside Ave. Lawrence KS 66045 [3]Department of Chemistry, University of Michigan, Ann Arbor Michigan 48109-1055

Introduction

In Gram-negative bacteria, the concentric structures of outer membrane, cell wall, and cytoplasmic membrane protect the cell from extracellular threats. Of these, the outer membrane is a particularly formidable barrier because of the structure of the lipopolysaccharide (LPS) constitutes the outer membrane's outer leaflet [119]. Imbedded in each bacterium's outer membrane are about a hundred varieties of barrel-shaped proteins [120]. Though these outer membrane proteins (OMPs) have a diverse array of functions, the most common role for these proteins is in transporting molecules across the membrane—importing nutrients and exporting toxins.

Due to the intrinsic impermeability of the outer membrane, accessibility of OMPs from the outside has led to interest in targeting the extracellular face of OMPs for the development of novel antibiotics which inhibit OMP folding [121]. Developing molecular plugs for the pores of OMP barrels would allow for the manipulation of bacterial transport, either starving bacteria by preventing the consumption of valuable nutrients or poisoning by preventing the release of toxins.

In the service of bacterial warfare, bacteria have evolved protein toxin systems called bacteriocins that hijack OMPs to cross the impermeable outer membrane and kill a target cell.

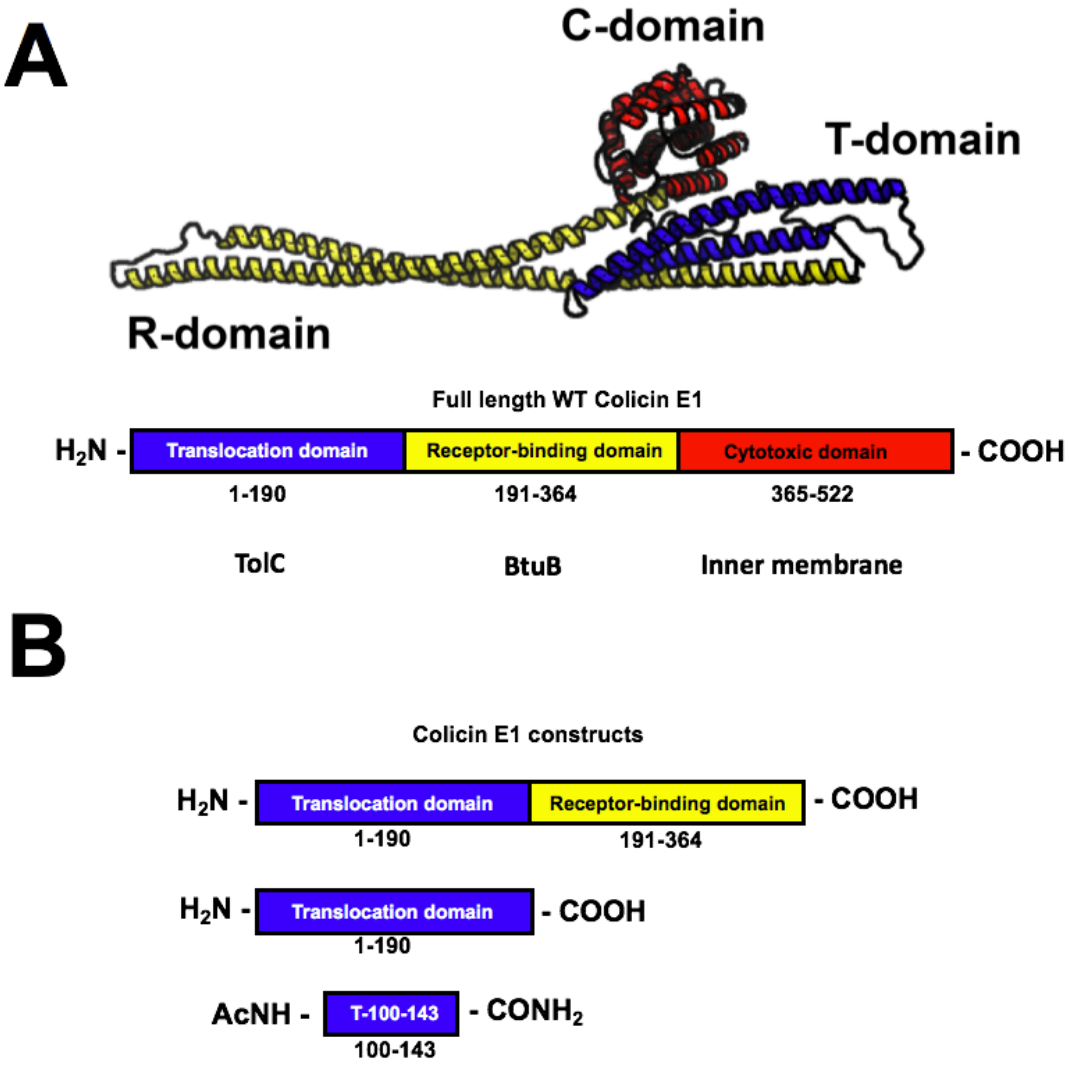


Figure 3.1: Colicin E1 domains (A) homology model of Colicin E1 using pore forming Colicin Ia as a template. Model generated using i-TASSER web-server https://zhanglab.ccmb.med.umich.edu/I-TASSER/ **(B)** N-terminal colE1 truncations used in this study

Colicins (*E. coli* specific bacteriocins) vary widely by receptor targets and killing mechanisms yet share a common triad domain architecture with an N-terminal translocation (T) domain, a receptor (R) binding domain, and a C-terminal cytotoxic (C) domain (**Figure 3.1A**). Colicin import is initialized by the binding of the center R-domain to an OMP target with high

affinity to localize and concentrate it onto the outer membrane. Once tethered to the outer membrane surface, the T-domain initiates translocation using a secondary OMP receptor distinct from the R-domain OMP target [32].

For some colicins the T and R domains remain in contact with their outer membrane while the cytotoxic domain kills the target cell [122, 123]. If T and R domains do not fully pass through their receptors, their binding events could be harnessed to disrupt native function of OMP receptors. Co-crystal structures of colicin E3 and E9 fragments bound to OmpF reveal that the barrel is obstructed by the peptides [124, 125].

Here we use the T and R domains of colicins to determine if the binding event between the N-terminal domains and the OMP pore is can plug and disrupt the native export function of a target OMP. Specifically we use the T and R-domain (**Figure 3.1B**) from colicin E1 which is known to bind to the antibiotic efflux protein TolC [126]. Assessing efflux inhibition of TolC allows us to explore a novel mechanism of antibiotic potentiation.

TolC is the outer membrane component of the acridine efflux pump which extrudes multiple classes of antibiotics such as erythromycin, chloramphenicol, tetracycline, doxorubicin, acriflavine [23, 24] as well as other compounds such as bile salts and detergents [22]. TolC is the efflux pump protein whose deletion makes E. coli most vulnerable to a wide variety of antibiotics [127]. Moreover, the pumps expression is linearly correlated to antibiotic resistance in clinical isolates [128]. As antibiotic resistance becomes a world-wide threat, finding proteins that can disable this primary mechanism of resistance could be useful therapeutics.

Colicin E1 uses TolC as the receptor of the T-domain and BtuB as the receptor of the R-domain [126]. Smaller constructs of the T-domain of colicin E1 have previously been shown to disrupt TolC channel conductance *in vitro* [33, 34]. Moreover, the n-terminal domain has been

shown to competitively inhibit cytotoxic activity of full length colicin E1. A stretch of residues 100-120 deemed the "TolC box" was determined to be required for this interaction with TolC. Structural characterization by circular dichroism of the colicin E1-TolC complex suggests colicin E1 inserts into TolC as a helical hairpin [34]. This *in vitro* structural characterization led to the proposal of a model where TolC structurally supports translocation rather than acting as a channel for translocation. If this is the case, the T and R domains of colicin E1 will not pass through but should remain within the luminal conduit of TolC effectively plugging the channel.

Here, through real-time efflux assays, synergistic MICs, and single molecule microscopy, we find that larger fragments of colicin E1 are able to inhibit TolC mediated efflux. We find that extracellular plugging of TolC reduces the amount of antibiotics required to inhibit cell growth—reducing this mode of antibiotic resistance. This work demonstrates that colicin E1 can be used to bacterial susceptibility to antibiotics and shows the potential for using colicin fragments for bacterial species-specific antibiotic potentiation and more broadly, species-specific blocking of OMP import/export function.

Results

Colicin E1 Localization

In order to determine colicin E1's utility as a plug for antibiotic efflux, we assessed if the colicin E1 constructs localizes on the cell surface or if it translocates across the outer membrane. We conducted an extracellular protease digestion assay. ColE1 stalling on the outer membrane surface would render it susceptible to digestion when trypsin is in the extracellular environment. Conversely, if the protein translocates into the intracellular environment, it would be shielded from digestion by the outer membrane. For the T-domain alone, there is no detectible binding.

This suggests the affinity to TolC is not strong enough to withstand the washing steps. For colE1-TR there is a band present in intact cells without trypsin showing binding to cells. When incubated with increasing amounts of trypsin, colE1-TR is digested at all trypsin concentrations indicating that it is localized to the outer membrane surface (**Figure 3.2A**) whereas the control, periplasmic chaperone SurA, was not degraded at any trypsin concentration [129].

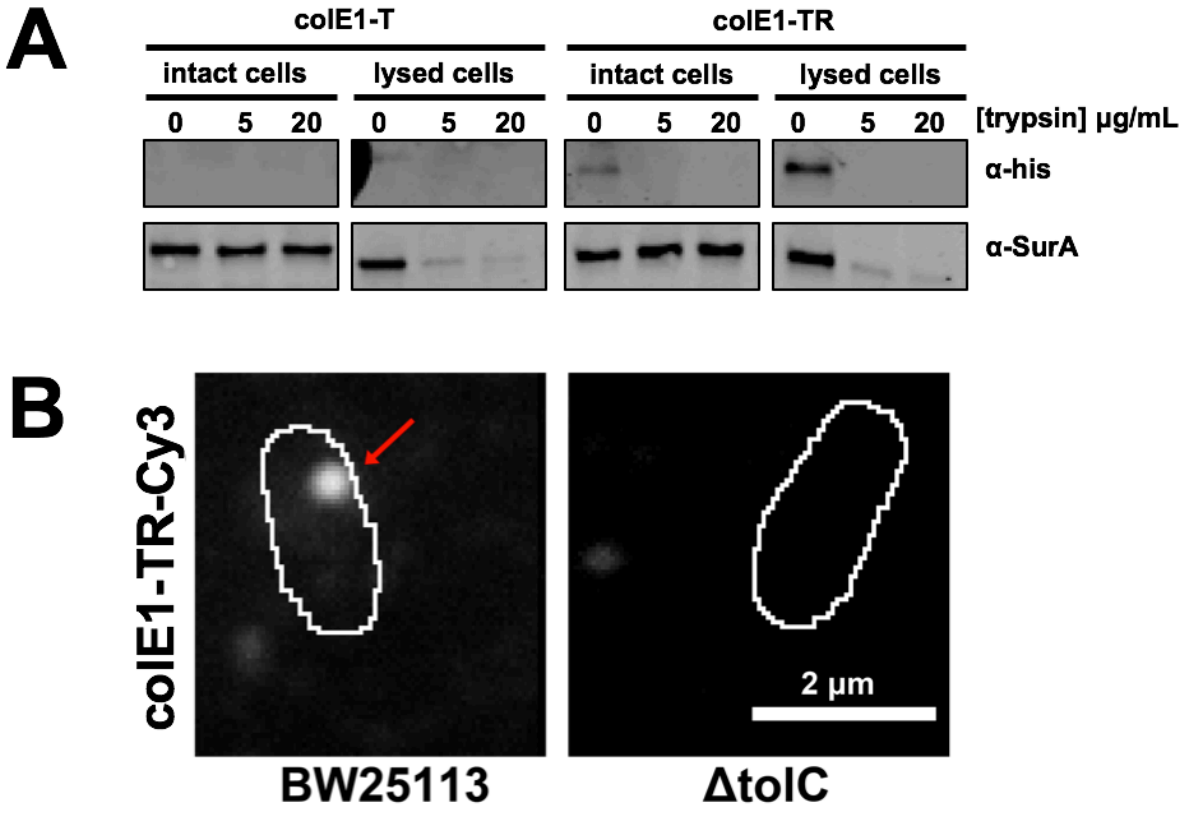


Figure 3.2: Colicin E1-TR localizes on the extracellular surface of *E. coli*: (A) Extracellular protease digest assay. SurA used as a periplasmic localization control. **(B)** Single-Molecule Microscopy: Red arrow indicates colE1-TR bound at ~12 molecules per punctum. Fluorescence images overlaid on outlines of cells from phase-contrast microscopy for BW25113 and Δ*tolC*. Scale bars: 2 µm.

After determining that colE1-TR remains at the cell surface, we probed binding and cell localization through single-molecule fluorescence microscopy. C-terminal cysteines were incorporated into colE1-T and colE1-TR to enable thiol coupling to the fluorescent dye cyanine 3

(Cy3). When colE1-TR is added to the extracellular environment, distinct punctate form on 94% (n=111) of cells containing TolC (**Figure 3.2B, left**). The appearance of puncta significantly decreases to 18% (n=99) of the population in cells lacking TolC (**Figure 3.2B, right**). When these puncta do appear, colE1-TR forms clusters of around a dozen molecules per punctum with a diameter of ~0.5 μm. The quantity of BtuB molecules and size of puncta is in agreement with previously determined clusters of BtuB [130, 131]. The existence of punctate formation but fewer puncta on cells lacking TolC is consistent with Colicin E1 initially binding to BtuB, but TolC enhancing the interaction.

As in the extracellular protease digestion, without the R domain, colE1-T shows no detectable binding to cells containing TolC or in a TolC knockout strain (**Figure A2.1**) indicating that the TolC:colE1-T interaction is much weaker than the BtuB:colE1-TR interaction. Moreover, when observing time courses of bound colE1-TR, all puncta remain immobile (**Figure A2.2**) further supporting that colE1-TR does not translocate [132] and that it binds to BtuB. Limited mobility has been previously described with fluorescently labeled BtuB which did not show FRAP on longtime scales [130]. Punctate localization is indicative of colE1-TR being associated with an OMP receptor since OMPs have been shown to exist in clustered islands in the OM thought to be mediated by protein-protein interactions due to molecular crowding effects in the membrane [130].

Colicin E1 Binding to TolC

The interaction of TolC and Colicin E1 fragments have previously been characterized *in vitro* by co-elution of the peptides by size exclusion chromatograph [34]. To ensure the interaction with our Colicin E1 constructs were specific for TolC, we assessed peak shifts for the

two colicin fragments when mixed with TolC. ColE1-T (**Figure 3.3A, right**) and TolC alone elute from the SEC at 14.8 mL and 10.8 mL, respectively. When mixed together the TolC peak shifts slightly to 10.7 and there is an increase in the peak intensity. Similarly, when TolC + colE1-TR are mixed (**Figure 3.3B, right**), a peak at 10.0 mL appears corresponding to the complex. Peak fractions from colE1-T (10.7 mL) and colE1-TR (10.0 mL) were analyzed by SDS- PAGE and shows the presence of both colicin E1 N-terminal constructs and TolC (**Figure 3.3A and B, left**).

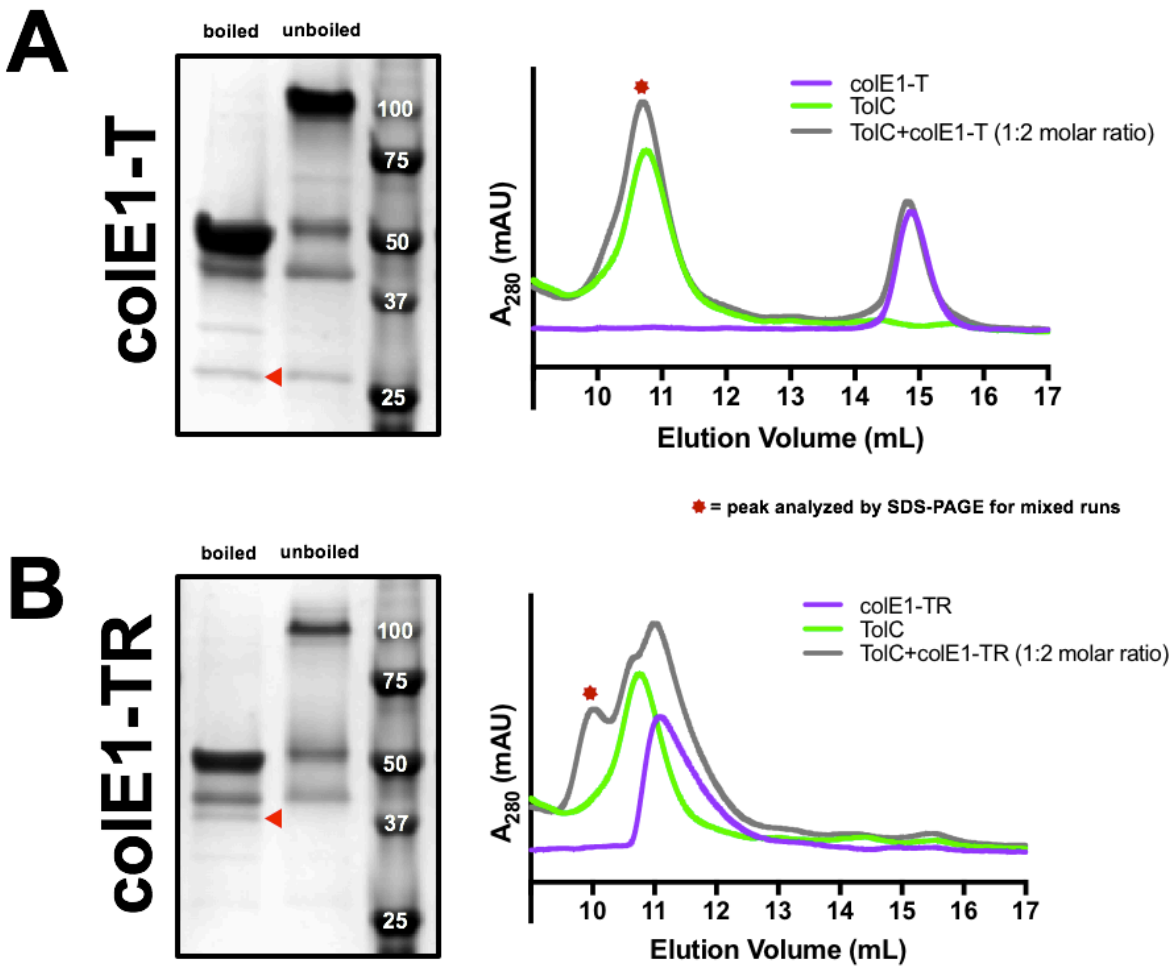


Figure 3.3: Colicin E1 fragments bind to TolC. Co-elution experiment to determine colE1 binding to TolC. **(A)** Co-elution of colE1-T *purple* line, TolC *green* line and colE1-T+TolC *grey* line. **(B)** Co-elution of colE1-TR *purple* line, TolC *green* line and colE1-TR+TolC *grey* line. The T domain did not show binding in our *in vivo* but binds to TolC *in vitro*. *Red stars* indicate fractions that were analyzed by SDS-PAGE. TolC runs as a trimer ~100 kDa in semi-native unboiled sample and ~50kDa as a monomer. *Red arrows* indicate the presence of colE1 constructs that have co-eluted with TolC

Colicin E1 Inhibits Active Efflux

Real-time efflux inhibition by colicin E1 fragments was assessed using a live cell assay using the NNN-dye which is effluxed by the acridine efflux pump [133] and which fluoresces when it is localized inside the cell. Efflux of NNN is turned off by the protonophore CCCP which neutralizes the proton motive force allowing for accumulation of the dye within the cell. Active efflux can then be monitored by the decay in fluorescent signal once proton motive is reenergized by the addition of glucose [133-137].

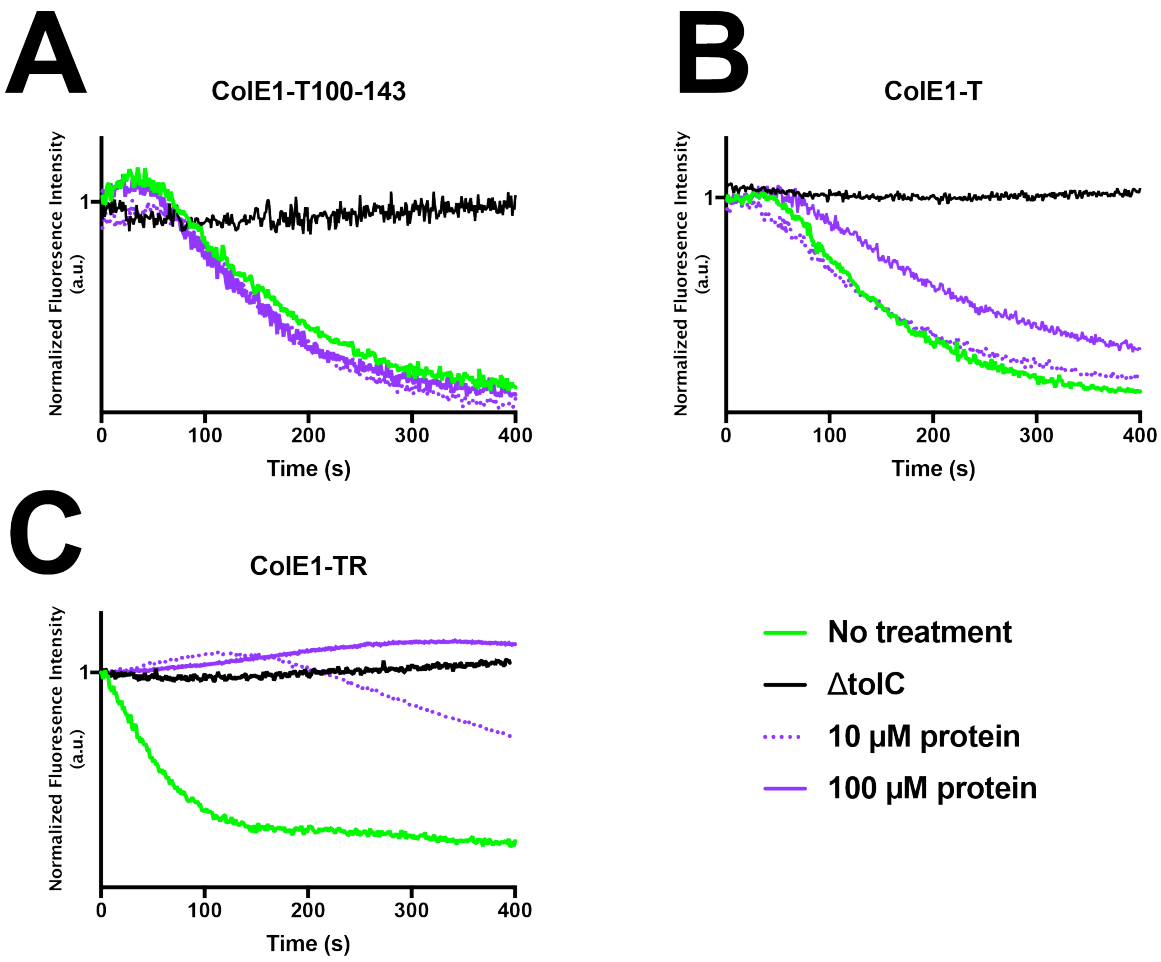


Figure 3.4: Effect of colE1 fragments on real-time efflux of N-(2-naphthyl)-1-naphthylamine (NNN) in *E. coli* BW25113 and BW25113Δ*tolC*: (A) An n-terminal peptide fragment composed of residues 100-143 encompassing the "TolC box" **(B)** The entirety of colE1-T domain **(C)** The entirety of the T and R domain. In all cases, BW25113 with no protein is shown in a *green* line. BW25113Δ*tolC* in *black* lines. BW25113 + 10 μM protein in purple dots. BW25113 + 100 μM purple lines. The TolC box peptide does not show activity against NNN efflux. ColE1-T moderately inhibits NNN efflux at 100 μM. ColE1-TR shows partial inhibition at 10 μM and full inhibition at 100 μM.

We assessed the effect of colicin E1 residues 100-143 (**Figure 3.1B**) on the real-time efflux of the fluorescent probe molecule NNN and found that efflux was not decreased by this peptide. **(Figure 3.4A**) though similar peptides have been found to bind TolC and occlude channel conductance [33, 34]. In contrast, the full colE1-T domain shows as a distinct decrease

in final fluorescence though it does not return to the baseline fluorescence indicating partial inhibition of efflux (**Figure 3.4B**). Finally, colE1-TR shows full inhibition of the acridine efflux pump as fluorescence is decreased after the addition of glucose. (**Figure 3.4C).**

Colicin E1 increases E. coli Susceptibility to Antibiotics

Because colE1-TR showed complete inhibition of efflux, we evaluated its ability to potentiate antibiotics through synergistic MICs. We chose a representative antibiotics from five different antibiotic-classes that are known TolC substrates: kanamycin, ciprofloxacin, erythromycin, rifampicin and the antimicrobial agent benzalkonium chloride. When 100 μM colE1-TR is used in combination with either kanamycin, ciprofloxacin, erythromycin or benzalkonium chloride, there is a significant reduction in the minimum inhibitory concentration (**Figure 3.5**). ColE1TR makes BW25113 *E. coli* ~2-4 fold more susceptible to these antibiotics (**Table 3.1**). Although there is a slight reduction in the MIC for rifampicin, the difference is less prominent.

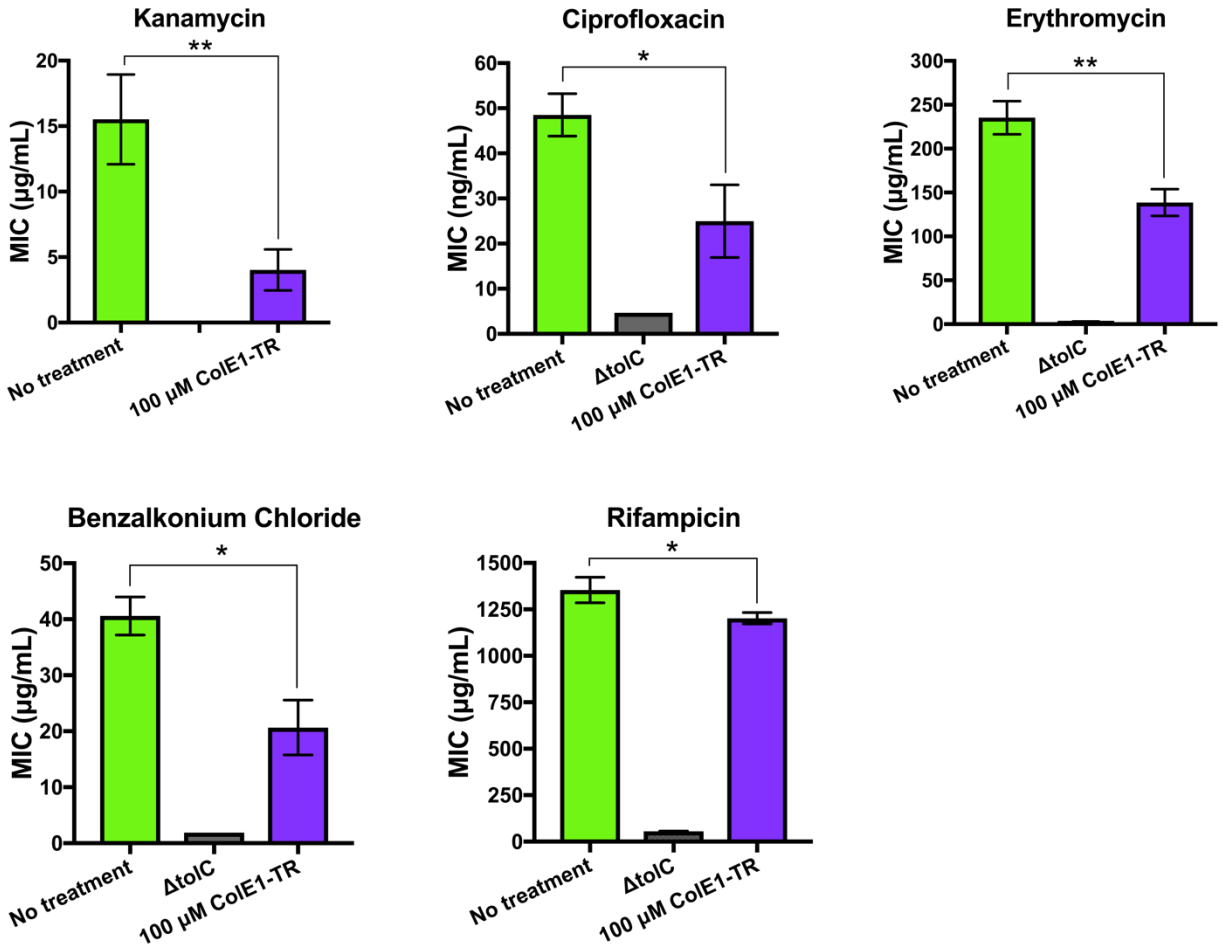


Figure 3.5: Colicin E1 TR potentiates antibiotics: Antibiotic susceptibility was determined in the absence (*green bars*) and presence (*purple bars*) of 100 µM colE1-TR for kanamycin, ciprofloxacin, erythromycin, benzalkonium chloride, and rifampicin in *E. coli* BW25113. MICs for BW25113ΔtolC (*grey bars*) are included as a reference for total loss of TolC. For kanamycin, the MIC for BW25113Δ*tolC* was not determined since the deletion of the *tolC* gene is accomplished by insertion of a kanamycin resistance cassette. All data for BW25113 were collected in triplicate with biological replicates, and error bars correspond to the standard error of the mean. Bars indicated have a statistically significant difference in their mean values (* with $p < 0.05$, ** with $p < 0.01$)

Table 3.1

Antibiotic	MIC BW25113	MIC BW25113 + colE1-TR	Fold reduction
Kanamycin	15.5 ±2.0 μg/mL	4.0 ±0.9 μg/mL	3.9x
Ciprofloxacin	48.5 ±2.0 ng/mL	25.0 ±4.7 ng/mL	1.9x
Erythromycin	253.3 ±10.8 μg/mL	138.7 ±10.8 μg/mL	1.8x
Benzalkonium Chloride	40.6 ±2.4 μg/mL	20.7 ±2.8 μg/mL	2.0x
Rifampicin	1354 ±39.5 μg/mL	1202 ±17.6 μg/mL	1.1x

Discussion

In this study we explore a novel method to modulate *E. coli* cell function by plugging an outer membrane protein channel using a fragment of a protein used for bacterial warfare. We leveraged the T domain of Colicin E1's natural ability to occlude TolC and sought to determine if it could also disrupt native function involved in efflux of antibiotics. We found that Colicin E1 lacking the cytotoxic domain, potentiates antibiotics of various classes including, aminoglycosides, fluoroquinolones, macrolides, and quaternary ammonium compounds.

Previous functional studies have identified the stretches of colicin E1 required for TolC binding leading to two proposed models for utilization of TolC. The first is the 'total thread' model in which the entire colicin is unfolded and TolC serves as a channel to thread through [138]. Structural studies of the binding interaction between TolC and N-terminal peptides of

colicin E1 have led to an alternative model for utilization of TolC. The second model proposes that colicin E1 inserts into TolC as a helical hairpin which serves as a buttress to facilitate cytotoxic domain entry [34, 138]. It is the intrinsic properties of the cytotoxic domain itself which allow it to cross the OM which may be mediated by anionic LPS. The results presented here support the latter model. Without the cytotoxic domain, we observed that colE1-TR remains stalled on the outer membrane exposed to the extracellular environment. The total thread model would allow cell entry of colE1-TR. As such, we propose a mechanism of efflux inhibition based on the Cramer "pillar model" of colicin E1 interaction with BtuB and TolC. (**Figure 3.6**) First, the R-domain binds to BtuB with high affinity and acts as an anchoring point for the colicin on the cell membrane. Second, the T-domain is able to search for TolC and insert into the channel stabilizing its association with the membrane and more importantly forming a plug that blocks exit of TolC substrates.

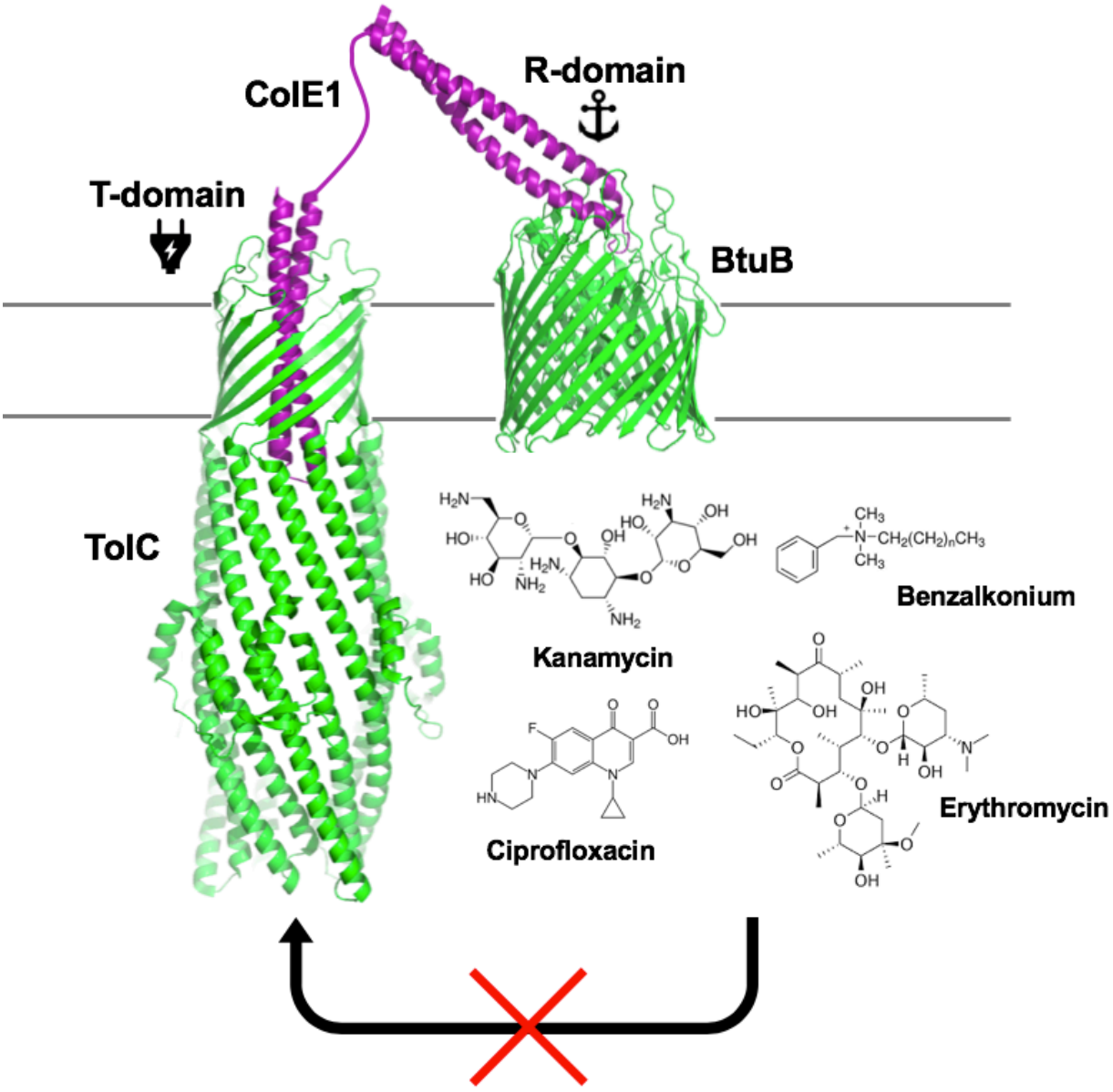


Figure 3.6: Model of colE1-TR inhibition of efflux: We propose a model for colE1-TR mediated inhibition of efflux adapted from the Cramer "pillar" model of interaction. The R-domain first binds to BtuB with high affinity and anchors colE1-TR to the surface of the cell. The T-domain then inserts into the lumen of TolC, plugging the channel and blocking exit of antibiotics. The crystal structure of BtuB bound to the R-domain of E3 was used to depict Crystal structure of BtuB and E3, homology model of Ia

We note that the truncation of colicin E1 required to inhibit efflux is much larger than the N-terminal fragments known to occlude TolC in channel conductance and cell protection. The

"TolC box" required for interaction with TolC was not sufficient to block efflux. The stretch of 43 residues may be too small to adopt the necessary secondary structure to fully occlude the channel or may be overpowered by the energy dependent process of efflux. The relative affinity of the T-domain for TolC is much weaker than the affinity of the R-domain to BtuB. Avidity by the additional interaction of the R-domain with BtuB is required for full inhibition of efflux. This is not surprising since evolution has produced colicin E1 for a specific purpose and thus the affinity is constrained to be only as tight as necessary. Its limited potency precludes it from any direct practical application. However, as a proof of concept, it offers enormous potential for further development. In principle, a more potent binder would not need the R-domain anchor. If the interaction between TolC and colE1-T can be resolved to atomic detail, the colicin T-domain would serve as a perfect candidate for optimization to engineer an effective antibiotic potentiator to prevent the spread of antibiotic resistance.

Methods

Expression and purification

Colicin E1 constructs

The gene for wild type Colicin E1-TR was synthesized as a double stranded linear fragment (Integrated DNA Technologies) and cloned into pET303 using megaprimer restriction free cloning. Deletion of the R-domain was conducted by inverse PCR to produce Colicin E1-T. Plasmids were transformed into E. coli BL21(DE3) cells and plated on LB + agar + 100 μg/mL carbenicillin. Single colonies were inoculated into 50 mL LB broth with 100 μg/mL carbenicillin and grown overnight at 37ºC with shaking at 250 RPM. Proteins were expressed by inoculating 1L of TB supplemented with 0.4% glycerol, 10 mM MgCl2 and 100 μg/mL carbenicillin with 20 mL of the overnight culture and grown at 37°C to an OD600 of 2.0 and induced with 1 mM IPTG. Expression cultures were grown at 15°C for 24 hours and harvested at 4,000g for 30 minutes at 4°C. Cell pellets were resuspended in (3 mL/g of cell pellet) lysis buffer (TBS, 5 mM MgCl2, 10 mM imidazole, 1mM PMSF, 10 μg/mL DNase, 0.25 mg/mL lysozyme) and lysed via sonication (2 minutes, 2s on, 8s off, 40% amplitude, QSonica Q500 with 12.7 mm probe) in an ice bath. Lysates were centrifuged at 4,000g for 10 minutes to remove un-lysed cells and debris. The supernatant was centrifuged again at 20,000 RPM in a Beckman Coulter J2-21 for 1 hour at 4°C. Clarified lysates were applied to a 5 mL HisTrap FF column and purified using an ÄKTA FPLC system with a 20 column volume wash step with (TBS, 25 mM imidazole) and eluted using a linear gradient from 0-50% elution buffer (TBS, 500 mM imidazole) in 10 column volumes. Proteins were concentrated in Amicon centrifugal filters with molecular weight cutoffs

of 10K and 30K for Colicin E1-T, Colicin E1-TR, respectively. Concentrated proteins were loaded onto a HiLoad Superdex 16/60 200 pg gel filtration column and eluted into PBS pH 7.4.

TolC

The gene for full length TolC was cloned into pTrcHis with a c-terminal 6x histidine tag. Plasmids were transformed into BL21(DE3)Omp8 and plated on LB + agar + 100 μg/mL carbenicillin. A single colony was picked and grown in LB-Lennox at 30ºC with shaking at 150 RPM overnight. In the morning, 1L of LB-Lennox was inoculated with 20 mL of the overnight culture and grown at 30ºC with shaking at 150 RPM until the culture reached an OD600 of 0.6 at which point protein expression was induced with 1mM IPTG for an additional 4 hours then harvested at 4,000g for 30 minutes at 4°C. Cell pellets were resuspended in 30 mL of Lysis buffer (TBS, 5 mM $MgCl_2$, 5 μg/mL DNase, 1mM PMSF) and lysed via sonication (2 minutes, 2s on, 8s off, 40% amplitude, QSonica Q500 with 12.7 mm probe) in an ice bath. Cell lysates were centrifuged at 15,600g for 30 seconds at 4ºC to remove un-lysed cells and inclusion bodies. Total membrane fractions were harvested by centrifuging the lysate at 15,600g for 30 minutes. The resulting pellet was resuspended in 2% n-lauryl-sarcosine and incubated at room temperature on a rotisserie for 30 minutes to solubilize the inner membrane. The total membrane fraction in 20 mM Tris pH 8, 400 mM NaCl, 2% n-lauryl-sarcosine was centrifuged at 15,600g for an additional 30 minutes with the resulting pellet containing the outer membrane fraction. The pellet containing the outer membrane fraction was resuspended in 20 mM Tris, 400 NaCl, 1.5% n-dodecyl-β-D-maltoside.

Single-Molecule Microscopy

Cysteine mutants for microscopy were purified as described above with the addition of 1 mM TCEP in all buffers to prevent covalent dimer formation. All subsequent steps were performed with limited exposure to light and the use of amber tubes. Cyanine3 maleimide was reconstituted in DMSO. Fluorophore labeling was achieved by mixing a 20-fold molar excess of Cyanine3 maleimide (Lumiprobe) to protein and incubating overnight at 4ºC. Free dye was removed by gel filtration on a Sephadex NAP-10 G-25 column. The sample was simultaneously buffer exchanged into storage buffer (PBS pH 7.4, 1 mM DTT, 1 mM EDTA). The degree of labeling was determined spectrophotometrically by the concentrations of the dye and protein using their respective extinction coefficients. (Cyanine3 ε_{548nm} = 162,000 L•mol^{-1}•cm^{-1}; colE1-T-E192C ε_{280nm} = 9, 970 L•mol^{-1}•cm^{-1}; colE1-TR-E366C ε_{280nm} = 14,440 L•mol^{-1}•cm^{-1}) Labeling efficiencies were ~75% and ~85% for colE1-T-E192C and colE1-TR-E366C, respectively. Protein concentrations were adjusted according to labeled protein.

Cultures of *E. coli* (BW25113 or Δ*tolC*) were grown in LB medium at 37ºC with shaking (180 rpm) overnight, then transferred to MOPS minimal medium with 0.2% glycerol and 1.32 mM K2HPO4, and grown at 37ºC for 13 h. A sample was transferred to MOPS medium and grown to turbidity at 37 ºC overnight. A 1-mL aliquot of culture was centrifuged for 2 min at 7,500 rpm to pellet the cells. The pellet was washed in 1 mL MOPS and centrifuged a second time. The supernatant was then removed, and the cell pellet was resuspended in 500 µL MOPS. A 1.0 µL droplet of concentrated cells was placed onto a glass slide. Then, 1.0 µL droplet of 1ug/mL Colicin E1-TR protein stock was added to the cells. The droplet was covered by an agarose pad (1% agarose in MOPS media) and a second coverslip.

Samples were imaged at room temperature using wide-field epifluorescence microscopy with sensitivity to detect single dye molecules as described previously [139]. Briefly, fluorescence was excited by a 561-nm laser (Coherent Sapphire 560-50) operated at a low power densities (4 – 5 μW), and imaged with an Olympus IX71 inverted microscope with a 100x, 1.40-NA oil-immersion objective and appropriate excitation, emission, and dichroic filters (Semrock LL01-561, Semrock BLP01-561, Semrock Di01-R561). A Photometrics Evolve electron multiplying charge-coupled device (EMCCD) camera with >90% quantum efficiency captured the images at a rate of 20 frames per second. Each detector pixel corresponds to a 49 x 49 nm area of the sample.

Extracellular Protease Digestion

BW25113 cells were grown to an OD_{600} of ~0.6. Cells were harvested by centrifugation at 4,000g for 5 minutes and resuspended in 1x Phosphate Buffered Saline (PBS). Cells were incubated with 10 μM protein and incubated for 1 hour at 37ºC with rocking. After incubation cells were harvested by centrifugation and washed 2x with PBS to remove any unbound protein. Cell pellets were resuspended in 5 mL of protease buffer (50 mM Tris pH=8, 7.5 mM $CaCl_2$) and OD_{600} normalized. Cultures were split into two samples for trypsin digestion 1) intact cells 2) lysed cells. For the lysed cell sample 0.25 mg/mL lysozyme was added and incubated at room temperature for 15 minutes. The cells were lysed by 5x freeze-thaw cycles by submerging in liquid nitrogen followed by thawing. For each cell condition (lysed and intact) the sample was further split into 6 aliquots. Aliquots were incubated with a final concentration of 0, 5, 20, 50, 100, 200 μg/mL trypsin. The reaction was incubated for 30 minutes with intermittent gentle

inverting of the tubes. After 30 minutes 100 mM PMSF was added to stop the digestion reaction. Samples were snap frozen in liquid nitrogen and stored at -20ºC until western blot analysis.

Co-elution

The interaction of TolC and colicin E1 fragments were determined by coelution on a size exclusion chromatography column. Purified TolC and colicin E1 fragments were mixed at a 1:2 molar ratio and incubated at room temperature for 1 hour before loading onto a Superdex 200 Increase 10/300 GL column (GE Healthcare). The protein was eluted with 1.5 column volumes into 20 mM Tris pH 8.0, 200 mM NaCl, 0.05% n-dodecyl-β-D-maltoside. Elution fractions were collected every 0.5 mL. Peak fractions were concentrated to 20 µL and analyzed by SDS-PAGE.

Real-time Efflux

A single colony of *E. coli* BW25113 was inoculated into 10 mL of LB and incubated overnight at 37°C. The next day, 50 mL of LB was inoculated with 1 mL of the overnight culture and grown to OD600 of ~0.8. Cells were pelleted at 3,500g for 5 minutes. Cells were resuspended to OD_{600} 1.5 in cold PBS with and without 10-100 µM Colicin proteins and incubated for 15 minutes on ice. To decouple proton motive force and turn off efflux, 100µM carbonyl cyanide m-chlorophenyl hydrazone (CCCP) was added. After an additional 15 minutes the efflux dye N-(2-naphthyl)-1-naphthylamine (NNN) was added to the cells to 10 µM. The cells were incubated at 25°C with shaking at 140 RPM for 2 hours. Cells were harvested at 3,500g for 5 minutes and washed once in 20 mM potassium phosphate buffer pH 7 with 1mM $MgCl_2$. Optical densities were adjusted to OD_{600} 1.0 and placed on ice. 2 mL of the cell suspension was loaded into a quartz cuvette (Agilent Technologies). Fluorescence was measured

with an Agilent Cary Eclipse fluorescence spectrophotometer with slit widths at 5 and 10 nm for excitation wavelength of 370 nm and emission wavelength of 415 nm. Fluorescence measurements were taken every 1 second. After 100 seconds 50 mM glucose was added to re-energize the cells and initiate efflux and fluorescence was collected for an additional 600 seconds.

Acknowledgments

We thank Dr. Mark Richter, Dwight Deay, and Priyanka Goyal for assistance with the fluorometer for efflux assays. We thank Johannes Thoma for providing the BL21(DE3)Omp8 strain for expression of TolC

CONCLUSION AND FUTURE DIRECTIONS

This dissertation has focused on studying what properties dictate protein-ligand interactions and the application of these principles to the design of novel ones. In chapter 1, the rules of how best to design an enzyme switch by chemical rescue of structure was explored by probing steric effects of protein-ligand interactions through structure activity relationship experiments, both from the perspective of the ligand and from the perspective of the protein [7]. The findings suggest that to best rescue an inactivating mutation that acts through a discrete conformational change, a small molecule ligand must replace the exact atoms removed by mutation. Additionally, the cavity can accommodate alternate larger ligands at the cost of the level of activity rescue. A smaller ligand that leaves a small void is favored to a larger molecule that introduces a steric class. With respect to tunability, we also found that alternate ligands can not only modulate the magnitude of activity but also shift selectivity towards a particular substrate.

One pitfall with using a single tryptophan to glycine mutation is that there is an intrinsic biophysical limit to the binding affinity for indole and molecules of this size. Ligand efficiencies and thus potencies seem to have hit a limit as seen with the requirement for millimolar concentrations of effectors necessary to illicit responses. At the end of chapter 1 we introduced the idea of adding a second mutation to accommodate an additional methyl group on the effector molecule which enhanced its binding affinity. In principle, this could be applied to introduce double or even triple mutants to find more potent molecules.

Moving forward, the generalizability of chemical rescue of structure will be greatly enhanced by the introduction of computational techniques. Double or triple residue mutations

can be exhaustively enumerated for a protein of interest by extracting constellations of atoms in 3D space associated with two or three residues. The next step would be to find the matching small molecule to replace the atoms removed by mutation. Using standard ligand based virtual screening methods like ROCS [98] introduced in chapter 2, the constellation of residues can be used as a query for a screen of a chemical library. The key in this approach would be to find coupled mutations that would not be *too* destabilizing. It may then be necessary to consider intermolecular mutation sites between protomers or intramolecular mutation sites between domains. This would allow the mutations to remain more conservative by only introducing one mutation per protomer or domain. Secondly, surface mutations are generally more tolerated than in buried residues. The frequency of successful applications of chemical rescue of structure based on indole rescue of a tryptophan to glycine mutation is surprisingly quite common. We anticipate that incorporating these developments will further expand its applicability.

Chapter 2 explored the conformational landscape of protein-protein interaction targets to find the druggable states. The flat and relatively featureless interfaces seen in the unbound crystal structures are not amenable to small molecule ligand binding. Here we model low energy natural pocket fluctuations that form on these interfaces through sampling by a biased simulation. For the instances where pockets do form, we applied a high-resolution high-throughput geometric shape comparison protocol to compare the pocket shapes that form to known pockets of previously discovered inhibitors. When comparing ensembles of the unbound protein and the ensembles of the protein with the molecule bound, we found that the pocket shapes sampled overlap. This indicates that proteins can adopt bound-like conformations without any information of the ligand. This is akin to the conformational selection model of binding where the surface fluctuations form *before* the ligand binds. Additionally, when all ensembles

were compared to each other, we found that the unbound protein can also sample pocket shapes that are distinct from any known inhibitor. These pockets reveal new untargeted states. The results of this work should allow for *de novo* design new small molecule inhibitors that are distinct from the current repertoire of available drugs.

A rigorous test of course is to design new small molecule ligands that can bind to these distinct states. What we have generated for the purpose of pocket shape comparisons in exemplars not only represents the pocket shape but can also be seen as an idealized ligand with perfect shape complementarity to the surface. In addition, other features that contribute to potent binding like hydrogen bond donors and acceptors can be included in the exemplar description. Analogous to the constellations being used as a query for identifying a ligand for the designed buried pocket, a virtual screen of exemplars can be used to find small molecules that match the pockets of PPI targets. The rapid speed of these algorithms allows screening of millions of compounds in a short amount of time. Another factor that must be taken into account that will require careful thought is curation of the chemical libraries. The ZINC15 database provides digital libraries of up to 750 million commercially available compounds that can be purchased for experimental validation [140]. Additionally, high resolution crystal or NMR structures will have to be determined to validate the models. With advances in computational methods such as these, previously undruggable targets are now becoming a tractable problem.

Lastly, in chapter 3 we investigated a novel way to target gram-negative bacteria by designing a molecular plug of a β-barrel protein on the surface of *E. coli*. Crossing the outer membrane of gram-negative bacteria to target intracellular targets has poses a difficult challenge in the discovery of new antibiotics. By plugging channels responsible for transportation nutrients or removal of toxins, we can essentially starve or poison the bacterial cell without the need to

cross the impermeable outer membrane. As a proof of concept, we investigated a naturally occurring protein that hijacks the efflux pump protein TolC to plug the channel. In its full form, colicin E1 utilizes TolC to deliver a cytotoxic domain to kill the cell in a process not fully understood [34, 138]. We found that a plug composed of an n-terminal fragment lacking the cytotoxic domain can block TolC. The molecular plug prevents active efflux and makes *E. coli* more susceptible to a diverse set of antibiotics.

The relatively low potency of the colicin E1 truncations we observed is not surprising given that evolution has constrained its sequence for a particular function unrelated to the purpose of plugging. However, as a platform for further development it offers huge potential for potentiating antibiotics. It will be imperative to determine the precise structural details of the colicin-TolC interactions in order to fully understand what properties are responsible for the interaction. To date, attempts to solve the co-crystal structure of colicin E1 and TolC have not yielded well diffracting crystals. Further crystal optimization is ongoing. Alternate methods include cryo-electron microscopy which has advanced to the point where atomic resolution is achievable. The tripartite complex of AcrAB-TolC was determined to 3.6Å resolution by cryo-EM [21]. However, the details of this interaction may be revealed, computational design methodologies could help guide design of a more potent inhibitor. Using algorithms like the design protocol within the Rosetta software suite [96] could potentially regenerate a sequence that would further optimize the packing interactions within the lumen of TolC. Another strategy could be to apply directed evolution to the currently existing sequence to achieve more potency without needing to know the structural details of the interaction. Of course, this will come with its own set of challenges such as a method to select and or screen for improved variants. Potential methods such as phage or ribosome display may prove useful. As is common with

many computationally designed proteins, it may be necessary sequentially couple the two by following a round of computational design with directed evolution. The design of a novel ligand to plug an outer membrane is unprecedented. We hope that the study presented in this dissertation will open the door for a new avenue of research to combat antibiotic resistance.

APPENDIX

Chapter 1 Supporting Information
A1.1 Supplementary Methods

Expression and purification of β-glycosidase

The gene for *Sulfolobus solfataricus* β-glycosidase, included in a pET29 vector, was transformed into *E. coli* Rosetta2(2DE3) pLysS cells and plated on LB + agar + 50 mg/mL kanamycin. Single colonies were inoculated into LB broth with 50 mg/mL kanamycin and grown to OD600 of 0.6 at 37°C. Protein expression was induced with 1mM IPTG for 4 hours at 37°C. Cells were centrifuged and resuspended in Lysis Buffer (50 mM sodium phosphate pH 6.8). The cells were lysed using three freeze thaw cycles in liquid nitrogen. The lysate was incubated at 75°C for 30 minutes to aggregate endogenous cellular proteins. The lysate was centrifuged at 27,000g for 30 minutes to clear cellular debris. The protein was purified from the supernatant by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 gel filtration column (GE Healthcare).

Functional assay

β-glycosidase kinetic assays were conducted at 37°C in black 96-well plates. 100 μL reaction volumes contained 58 nM β-glycosidase in Assay Buffer (50 mM sodium phosphate pH 6.5), and effector concentrations ranging from 0-5 mM in 10% DMSO, and 750 μM of FDGal / FDGlc substrate dissolved in 0.1% ethanol, 0.1% DMSO, and 99.8% water. Product formation was monitored by accumulation of fluorescein using a Synergy 2 Microplate reader (Biotek), with an excitation filter of 485 nm and an emission filter of 528 nm. Initial velocities were

calculated by linear regression in GraphPad Prism Version 6.0. Fitting to the simplified allosteric rate equation (**Equation 1.1**) was carried out using GraphPad Prism Version 6.0, to obtain dissociation constants of the various effectors.

Crystal Structure of holo 5-fluoroindole W33G

Crystallization and Data Collection:

A β-glycosidase construct containing the W33G mutation concentrated to 10.0 mg/mL in Hepes pH 7.0 was screened for crystallization in Compact Jr. (Emerald biosystems) sitting drop vapor diffusion plates. Crystals were obtained from the Index screen condition E1 (Hampton Research, 45% 2-methyl-2,4-pentanediol, 100 mM Bis-Tris pH 6.5, 200 mM $CaCl_2$) using equal volumes of protein and crystallization solution equilibrated against 100uL of the latter at 20°C. Prismatic crystals were obtained within 24 to 48 hours. To prepare the 5-fluoroindole bound complex, single crystals were transferred to a drop containing 100 mM indole dissolved in crystallization solution and incubated for 48 hours. Crystals were harvested directly from the soak solution and stored in liquid nitrogen for data collection. Data were collected at the Advanced Photon Source IMCA-CAT beamline 17ID using a Dectris Pilatus 6M pixel array detector.

Structure Solution and Refinement:

Intensities were integrated and scaled using XDS [141] and Aimless [142] respectively. Structure solution was carried out by molecular replacement with Phaser [143] using the isomorphous non-crystallographic dimer of a previously determined indole bound structure of the W33G mutant (PDB: 4EAN [5]) as the search model. Structure refinement and manual model building were performed with Phenix [42] and Coot [144] respectively. Structure validation was carried out using Molprobity [42]. Disordered side chains were truncated to the

point where difference electron density could be observed. Relevant crystallographic data are provided in **Table A1.3**.

A1.2 Supplementary Results

Large difference density (Fo-Fc) greater than 3 and consistent with 5-fluoroindole was observed near Gly33 of each subunit and was modeled in the orientation as shown in **Figures A1.4a and A1.4b**. Based on the difference electron density maps, the 5-fluoroindole molecules seem to fit best to the electron density in the orientation as modeled. In addition, feature-enhanced (FEM) [113] electron density maps were computed which are depicted in **Figures A1.4c and A1.4d**. Combined with the observed electron density from the Fo-Fc omit maps, the orientation that was modeled for the 5-fluoroindole molecules seems to be the most plausible. However, it is feasible that a minor population exists such the molecules adopt an alternate orientation that does not contribute to the observed electron density.

Superposition of the 5-fluoroindole bound model with the indole bound (PDB: 4EAN) using GESAMT [145] indicated that the structures are nearly identical with RMSD deviations of 0.17 Å and 0.18 Å between Cα-atoms of subunits A and B respectively (483 and 485 residues). The superimposed non-crystallographic symmetry (NCS) related dimers are shown in **Figure A1.5a**. Given the high degree of similarity, it was not surprising that Trp 433 adopts the “ligand bound” conformation in which the side chain is moved away from the indole binding region as shown in **Figures A1.5b and A1.5c**.

A1.3 Supplemental Tables

	Peak Integral (normalized to WT)
WT + 6FI	1.00
W33G + 6FI	0.48
W33G + 6FI + indole	0.99
W33G + 6FI + 3-methylindole	0.90
W33G + 6FI + N-methylaniline	0.90
W33G + 6FI + 2-methylindole	0.55

Table A1.1: ^{19}F NMR peak integrals. Integrated peak intensities are directly proportional to the number of ^{19}F nuclei giving the signal, and are therefore reported relative to the sample containing wild type β-glycosidase.

	LE (kcal/mol·atom)
indole (W33G)	-0.31
3-methylindole (W33G)	-0.31
N-methylaniline (W33G)	-0.32
5-fluoroindole (W33G)	-0.42
N-methylaniline (W33A)	-0.27
indole (W33A)	-0.24
3-methylindole (W33A)	-0.28
indole (W33G_V37A)	-0.33
5-methylindole (W33G_V37A)	-0.43

Table A1.2: Ligand efficiency (binding free energy per heavy atom) for each ligand/protein pair reported in this study. Binding free energy is calculated from *Kd* values obtained by fitting enzyme kinetics.

	β-gly W33G + 5-fluoroindole
Data Collection	
Unit-cell parameters (Å, º)	*a*=b=167.71, *c*=95.87
Space group	$P3_121$
Resolution (Å)♯	48.41-1.75 (1.78-1.75)
Wavelength (Å)	1.0000
Temperature (K)	100
Observed reflections	1,526,719
Unique reflections	155,679
<I/(I)>♯	10.7 (1.9)
Completeness (%)♯	100 (100)
Multiplicity♯	9.8 (10.1)
R_{merge} (%)♯,	11.5 (116.2)
R_{meas} (%)♯, ✓	12.2 (122.4)
R_{pim} (%)♯, ✓	3.9 (38.4)
$CC_{1/2}$ ♯,	0.998 (0.821)
Refinement	
Resolution (Å)	34.06-1.75
Reflections (working/test)	147,863 / 7,720
R_{factor} / R_{free} (%)	15.3 / 17.3
No. of atoms (protein / 5-fluoroindole / water)	7,981 / 20 / 568
Model Quality	
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (º)	1.005
Average *B* factor (Å^2)	
All Atoms	29.8
Protein	29.1
Indole	46.3
Water	38.7
Coordinate error (maximum likelihood) (Å)	0.17
Ramachandran Plot	
Favored (%)	97.8
Allowed (%)	2.0

♯) Values in parenthesis are for the highest resolution shell.

) $R_{merge} = \sum_{hkl}\sum_{i} |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_{i} I_i(hkl)$, where $I_i(hkl)$ is the intensity measured for the *i*th reflection and $\langle I(hkl)\rangle$ is the average intensity of all reflections with indices hkl.

) $R_{factor} = \sum_{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \sum_{hkl} |F_{obs}(hkl)|$; Rfree is calculated in an identical manner using 5% of randomly selected reflections that were not included in the refinement.

✓) R_{meas} = redundancy-independent (multiplicity-weighted) R_{merge}[142, 146]. R_{pim} = precision-indicating (multiplicity-weighted) R_{merge}[147, 148].

) $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data [149, 150].

Table A1.3. Crystallographic data for β-glycosidase W33G with 5-fluoroindole bound, refined to 1.75 Å resolution.

A1.4 Supplemental Figures

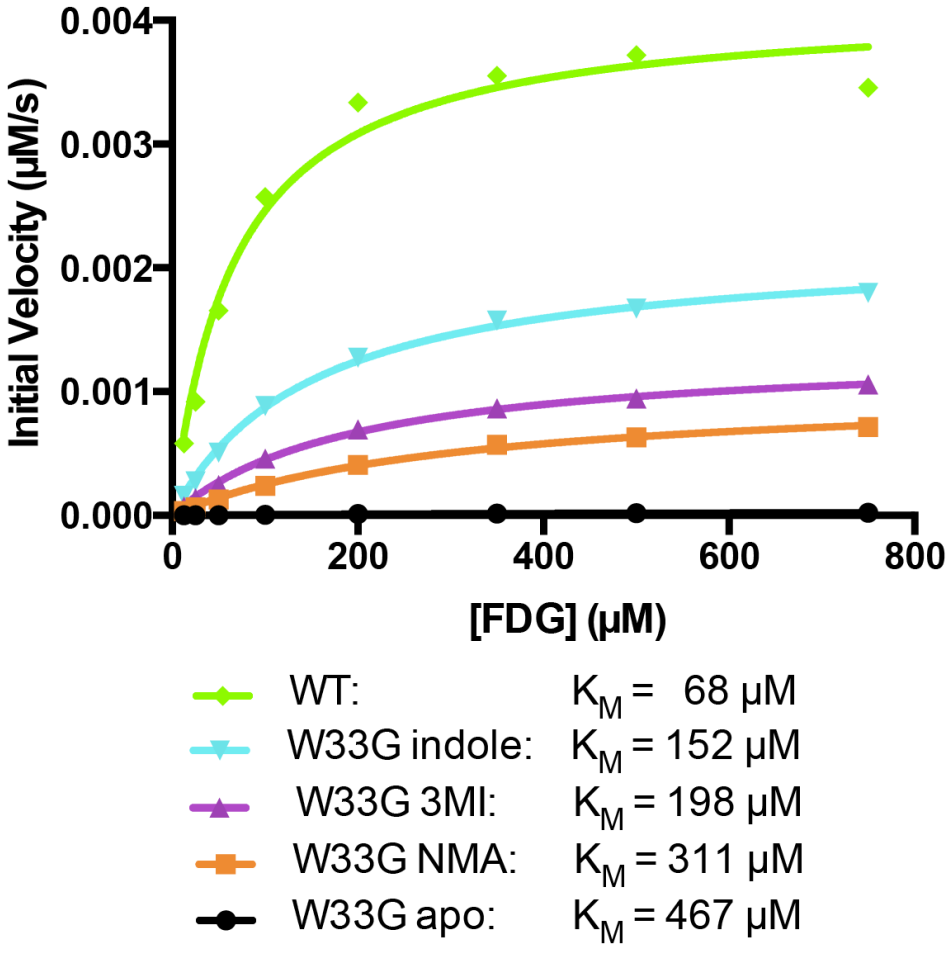


Figure A1.1: Initial velocities as a function of substrate concentration (FDGal, fluorescein di-β-D-galactopyranoside), in the presence of 5 mM of various effectors (indole, 3-methylindole, and N-methylaniline). Data for each curve were fit to the Michaelis-Menton equation, to obtain the apparent K_M values under each condition. In each case the K_M values are below the 750 μM substrate concentration used in each of the other experiments reported in this study. K_M values were not explicitly measured for all of the effector ligands used in this study, nor were they determined for protein mutants other than W33G.

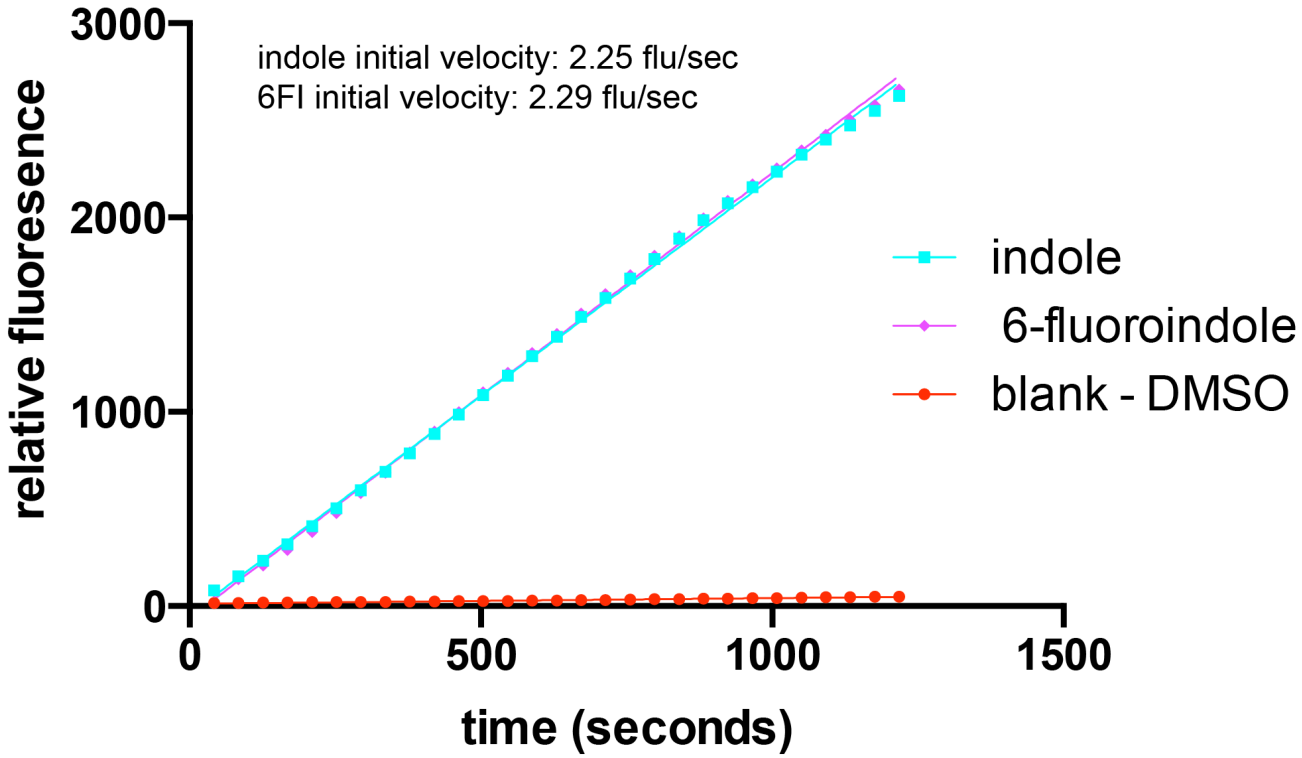


Figure A1.2: 6-fluoroindole rescues β-gly W33G to the same extent as indole itself. Product formation was measured by fluorescence using 750 μM fluorescein di-β-galactopyranoside (FDGal) as substrate.

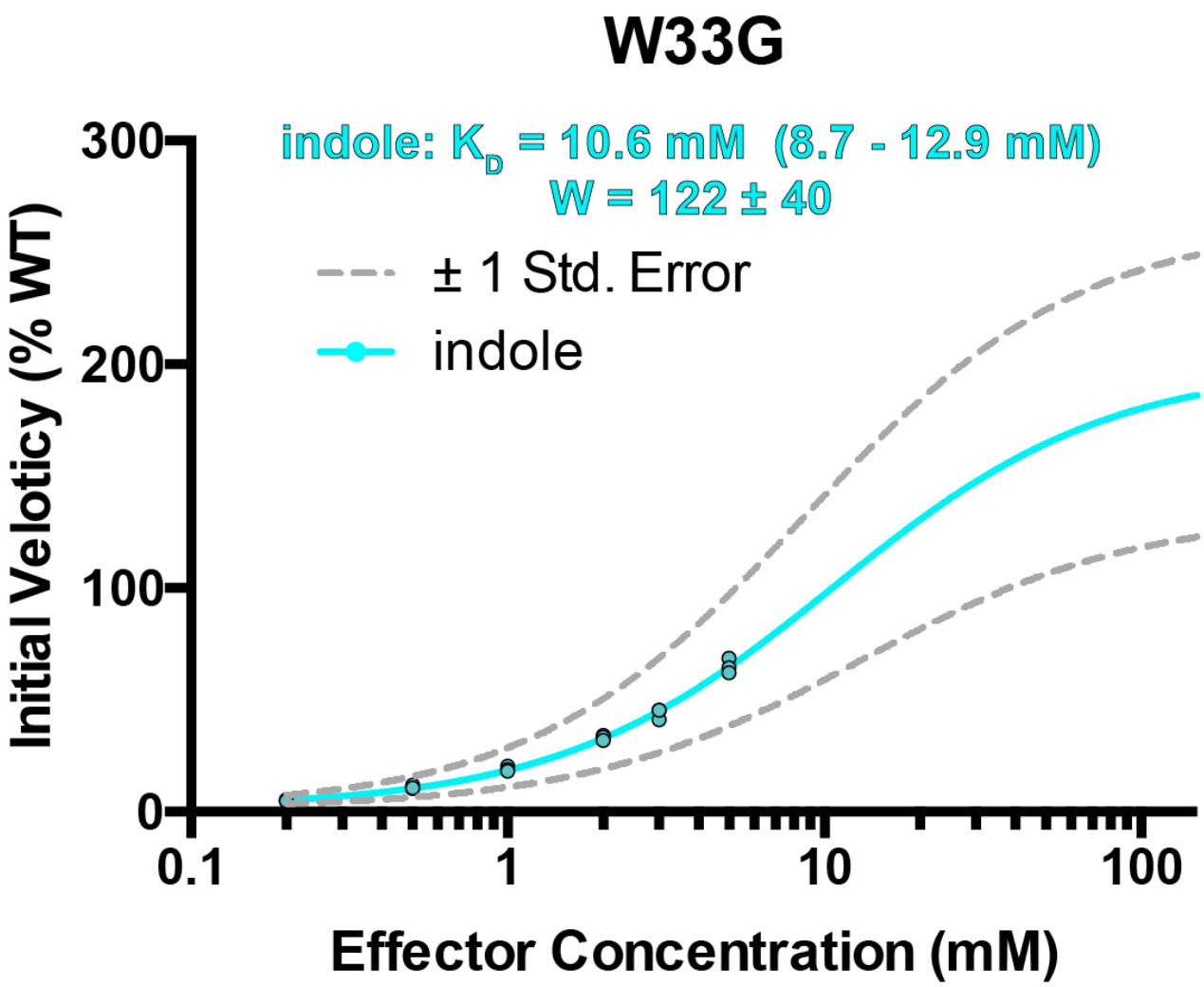


Figure A1.3: Maximal upper and lower bounds of the initial velocity. The range corresponding to one standard error above and below the best-fit values of K_D and W are reported. The upper dashed gray line corresponds to the initial velocity that would be observed from K_D = 8.7 mM and W = 162 (the lowest K_D value with the highest W value). The lower dashed gray line corresponds to the initial velocity that would be observed from K_D = 12.9 mM and W = 82 (the highest K_D value with the lowest W value). Thus, the dashed gray lines represent the maximal confidence intervals of the curve, at ± 1 standard error.

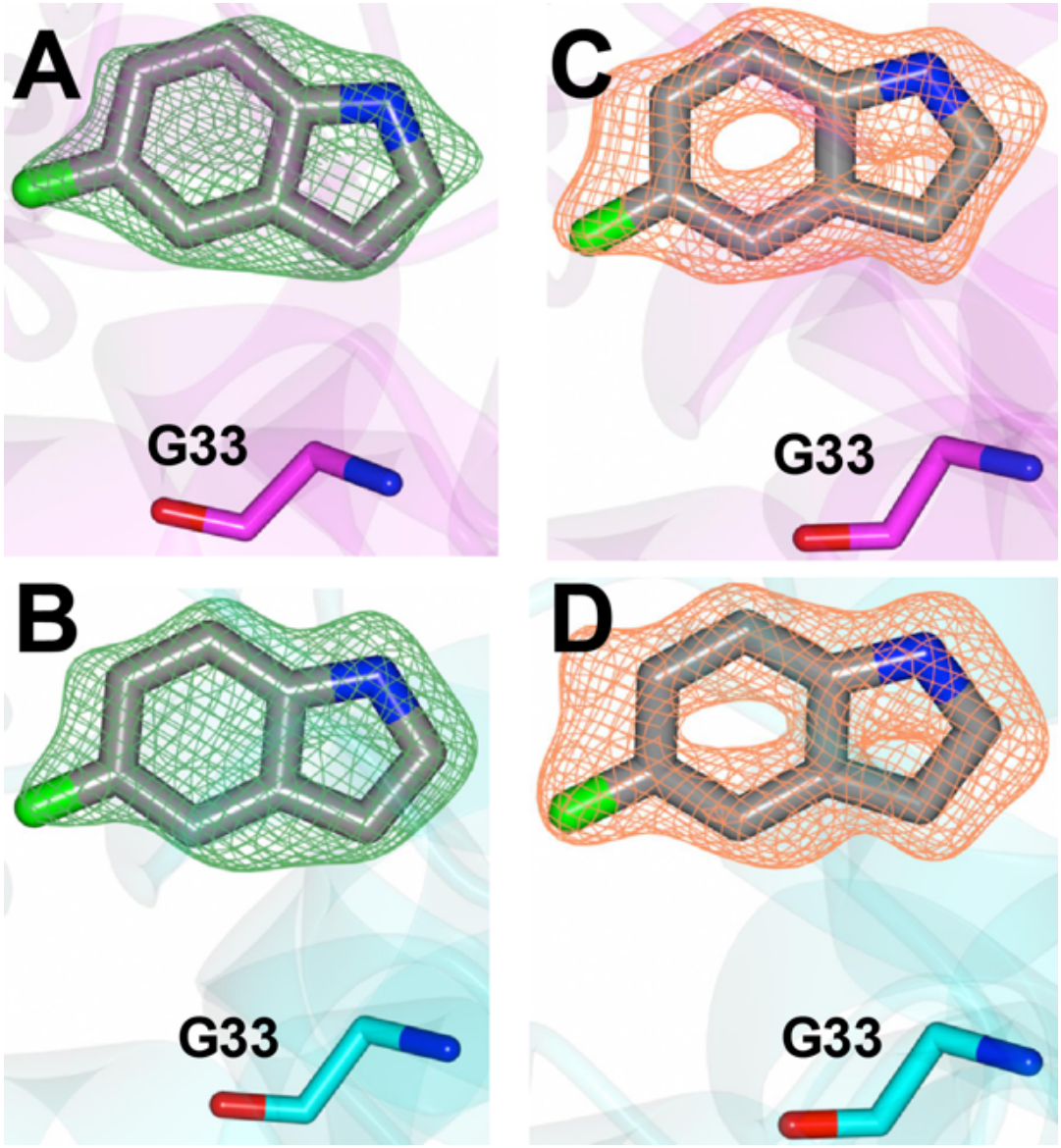


Figure A1.4: Electron density maps showing the orientation of the 5-fluoroindole molecules within the β-gly subunits A (*magenta*) and B (*cyan*). Panels **(A)** and **(B)**: Fo-Fc omit maps (*green mesh*) contoured at 3σ. Panels **(C)** and **(D)**: Feature enhanced 2Fo-Fc maps (*tan mesh*) contoured at 1.5σ.

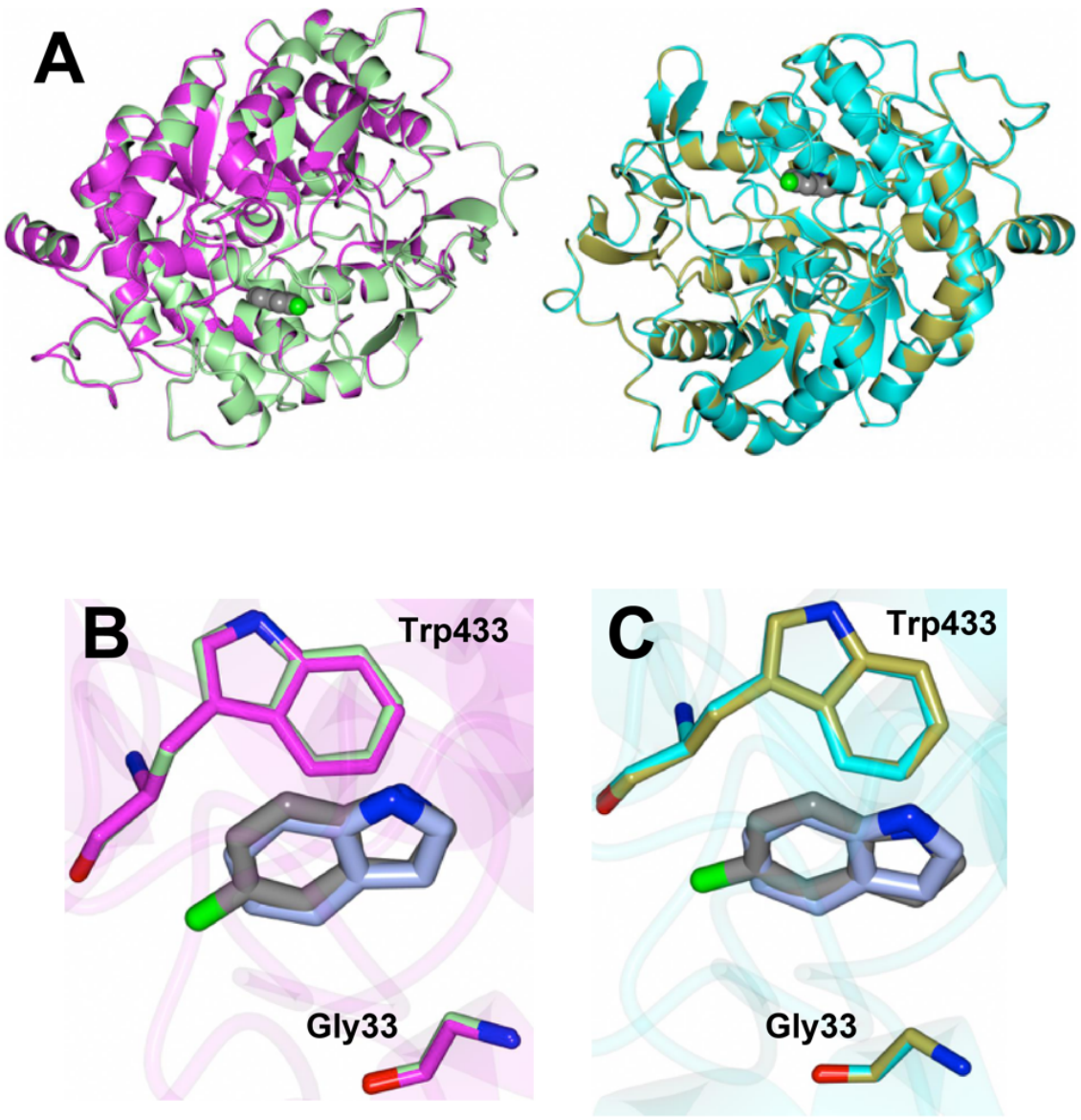


Figure A1.5: Superposition of 5-fluoroindole-bound and indole-bound β-gly structures. **(A)** 5-fluoroindole bound β-gly NCS dimer showing subunits A (*magenta*) and B (*cyan*) superimposed with subunits A (*light green*) and B (*gold*) of indole bound β-gly (PDB: 4EAN). 5-fluoroindole molecules are shown as spheres. **(B)** and **(C)**: Zoomed in view of the indole binding regions. 5-fluoroindole and indole molecules are rendered as *gray* and *light blue* respectively.

Appendix A.2

Chapter 3 Supporting Information

A2.1 Supplemental Figures

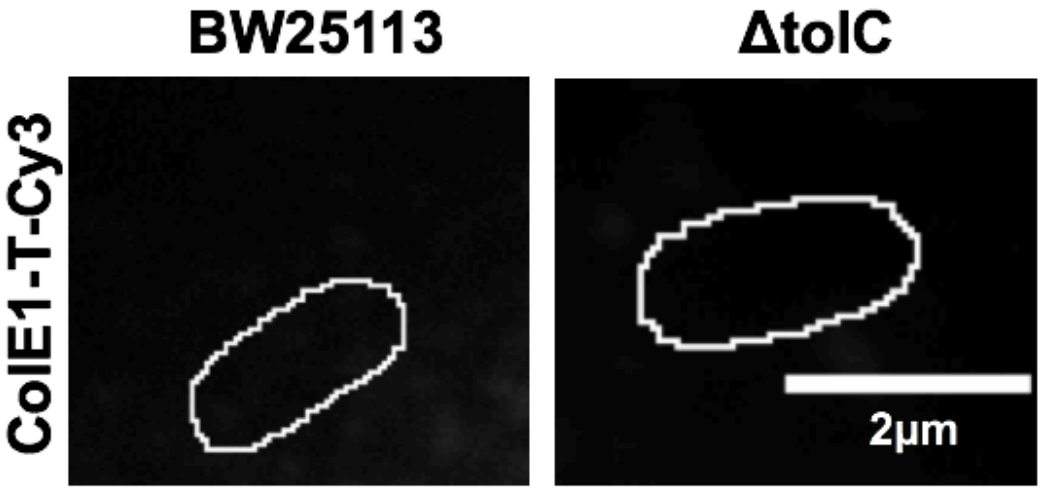


Figure A2.1: ColE1-T-Cy3 does not bind to cells.

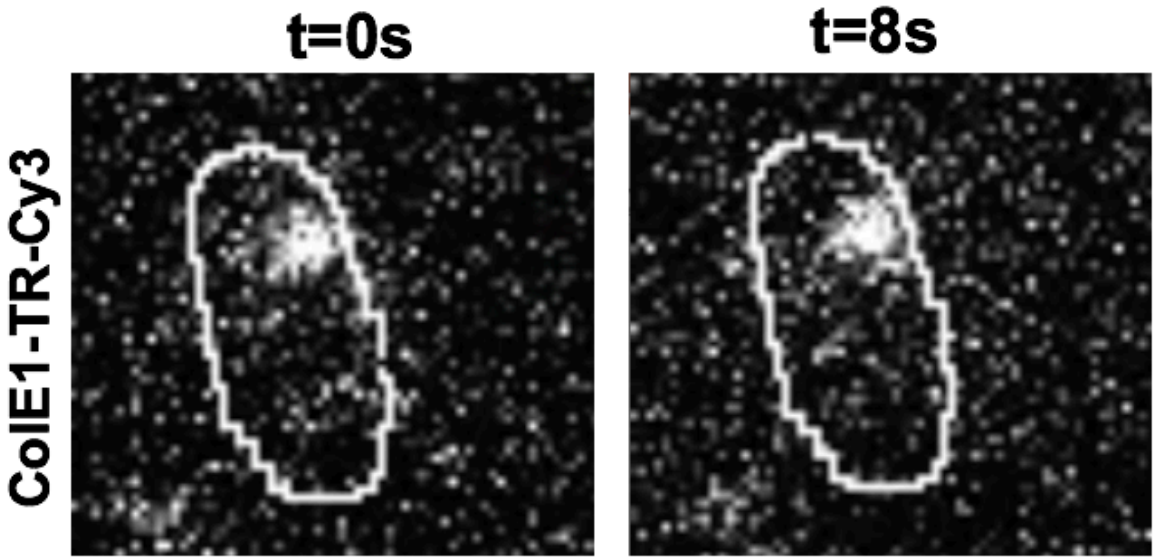


Figure A2.2: Time course of ColE1-TR-Cy3 bound to *E. coli* BW25113. The puncta formed remains immobile.

REFERENCES

1. Karginov, A.V., et al., *Engineered allosteric activation of kinases in living cells.* Nat Biotechnol, 2010. **28**(7): p. 743-U1756.
2. Banaszynski, L.A., et al., *A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules.* Cell, 2006. **126**(5): p. 995-1004.
3. Stein, V. and K. Alexandrov, *Synthetic protein switches: design principles and applications.* Trends Biotechnol, 2015. **33**(2): p. 101-110.
4. Ke, W., et al., *Structure of an engineered beta-lactamase maltose binding protein fusion protein: insights into heterotropic allosteric regulation.* PLoS One, 2012. **7**(6): p. e39168.
5. Deckert, K., et al., *Designing allosteric control into enzymes by chemical rescue of structure.* J Am Chem Soc, 2012. **134**(24): p. 10055-60.
6. Xia, Y., et al., *The designability of protein switches by chemical rescue of structure: mechanisms of inactivation and reactivation.* J Am Chem Soc, 2013. **135**(50): p. 18840-9.
7. Budiardjo, S.J., et al., *Full and Partial Agonism of a Designed Enzyme Switch.* ACS Synth Biol, 2016. **5**(12): p. 1475-1484.
8. Kaiser, C.E., et al., *Modulating Antibody Structure and Function through Directed Mutations and Chemical Rescue.* ACS Synth Biol, 2018. **7**(4): p. 1152-1162.
9. Gowthaman, R., E.J. Deeds, and J. Karanicolas, *Structural properties of non-traditional drug targets present new challenges for virtual screening.* J Chem Inf Model, 2013. **53**(8): p. 2073-81.

10. Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface.* Science, 1995. **267**(5196): p. 383-6.

11. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces.* J Mol Biol, 1998. **280**(1): p. 1-9.

12. Ghanakota, P. and H.A. Carlson, *Driving Structure-Based Drug Discovery through Cosolvent Molecular Dynamics.* J Med Chem, 2016. **59**(23): p. 10383-10399.

13. Bowman, G.R. and P.L. Geissler, *Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites.* Proc Natl Acad Sci U S A, 2012. **109**(29): p. 11681-6.

14. Johnson, D.K. and J. Karanicolas, *Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface.* PLoS Comput Biol, 2013. **9**(3): p. e1002951.

15. Ventola, C.L., *The antibiotic resistance crisis: part 1: causes and threats.* P T, 2015. **40**(4): p. 277-83.

16. Pendleton, J.N., S.P. Gorman, and B.F. Gilmore, *Clinical relevance of the ESKAPE pathogens.* Expert Rev Anti Infect Ther, 2013. **11**(3): p. 297-308.

17. Smith, P.A., et al., *Optimized arylomycins are a new class of Gram-negative antibiotics.* Nature, 2018. **561**(7722): p. 189-194.

18. Blair, J.M., et al., *Molecular mechanisms of antibiotic resistance.* Nat Rev Microbiol, 2015. **13**(1): p. 42-51.

19. Du, D., et al., *Multidrug efflux pumps: structure, function and regulation.* Nat Rev Microbiol, 2018. **16**(9): p. 523-539.

20. Du, D., et al., *Structure of the AcrAB-TolC multidrug efflux pump.* Nature, 2014. **509**(7501): p. 512-5.

21. Wang, Z., et al., *An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump.* Elife, 2017. **6**.

22. Anes, J., et al., *The ins and outs of RND efflux pumps in Escherichia coli.* Front Microbiol, 2015. **6**: p. 587.

23. Yamaguchi, A., R. Nakashima, and K. Sakurai, *Structural basis of RND-type multidrug exporters.* Front Microbiol, 2015. **6**: p. 327.

24. Ramaswamy, V.K., et al., *Molecular Rationale behind the Differential Substrate Specificity of Bacterial RND Multi-Drug Transporters.* Sci Rep, 2017. **7**(1): p. 8075.

25. Ayhan, D.H., et al., *Sequence-Specific Targeting of Bacterial Resistance Genes Increases Antibiotic Efficacy.* PLoS Biol, 2016. **14**(9): p. e1002552.

26. Farha, M.A., et al., *Collapsing the proton motive force to identify synergistic combinations against Staphylococcus aureus.* Chem Biol, 2013. **20**(9): p. 1168-78.

27. Hards, K. and G.M. Cook, *Targeting bacterial energetics to produce new antimicrobials.* Drug Resist Updat, 2018. **36**: p. 1-12.

28. Abdali, N., et al., *Reviving Antibiotics: Efflux Pump Inhibitors That Interact with AcrA, a Membrane Fusion Protein of the AcrAB-TolC Multidrug Efflux Pump.* ACS Infect Dis, 2017. **3**(1): p. 89-98.

29. Haynes, K.M., et al., *Identification and Structure-Activity Relationships of Novel Compounds that Potentiate the Activities of Antibiotics in Escherichia coli.* J Med Chem, 2017. **60**(14): p. 6205-6219.

30. Higgins, M.K., et al., *Structure of the ligand-blocked periplasmic entrance of the bacterial multidrug efflux protein TolC.* J Mol Biol, 2004. **342**(3): p. 697-702.

31. Gilardi, A., et al., *Biophysical characterization of E. coli TolC interaction with the known blocker hexaamminecobalt.* Biochim Biophys Acta Gen Subj, 2017. **1861**(11 Pt A): p. 2702-2709.

32. Cascales, E., et al., *Colicin Biology.* Microbiology and Molecular Biology Reviews, 2007. **71**(1): p. 158-229.

33. Zakharov, S.D., et al., *Colicin occlusion of OmpF and TolC channels: outer membrane translocons for colicin import.* Biophys J, 2004. **87**(6): p. 3901-11.

34. Zakharov, S.D., X.S. Wang, and W.A. Cramer, *The Colicin E1 TolC-Binding Conformer: Pillar or Pore Function of TolC in Colicin Import?* Biochemistry, 2016. **55**(36): p. 5084-94.

35. Czlapinski, J.L., et al., *Conditional glycosylation in eukaryotic cells using a biocompatible chemical inducer of dimerization.* J Am Chem Soc, 2008. **130**(40): p. 13186-13187.

36. Pratt, M.R., E.C. Schwartz, and T.W. Muir, *Small-molecule-mediated rescue of protein function by an inducible proteolytic shunt.* Proc Natl Acad Sci U S A, 2007. **104**(27): p. 11209-14.

37. Golynskiy, M.V., et al., *Engineering protein switches: sensors, regulators, and spare parts for biology and biotechnology.* Chembiochem, 2011. **12**(3): p. 353-61.

38. Merkx, M., et al., *Rational design of FRET sensor proteins based on mutually exclusive domain interactions.* Biochem Soc Trans, 2013. **41**: p. 1201-1205.

39. Buskirk, A.R., et al., *Directed evolution of ligand dependence: small-molecule-activated protein splicing.* Proc Natl Acad Sci U S A, 2004. **101**(29): p. 10505-10.

40. Lockless, S.W. and T.W. Muir, *Traceless protein splicing utilizing evolved split inteins.* Proc Natl Acad Sci U S A, 2009. **106**(27): p. 10999-1004.

41. Hiraga, K., et al., *Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center.* J Mol Biol, 2009. **393**(5): p. 1106-17.

42. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

43. Guntas, G., et al., *Directed evolution of protein switches and their application to the creation of ligand-binding proteins.* Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11224-9.

44. Choi, J.H., T. Xiong, and M. Ostermeier, *The interplay between effector binding and allostery in an engineered protein switch.* Protein Sci, 2016: p. [Epub ahead of print].

45. Ang, J., et al., *Tuning Response Curves for Synthetic Biology.* ACS Synth Biol, 2013. **2**(10): p. 547-567.

46. Murphy, K.F., et al., *Tuning and controlling gene expression noise in synthetic gene networks.* Nucleic Acids Res, 2010. **38**(8): p. 2712-26.

47. Granier, S. and B. Kobilka, *A new era of GPCR structural and chemical biology.* Nat Chem Biol, 2012. **8**(8): p. 670-3.

48. Abbas, A. and B.L. Roth, *Arresting serotonin.* Proc Natl Acad Sci U S A, 2008. **105**(3): p. 831-832.

49. Srinivasan, S., et al., *Ligand-binding dynamics rewire cellular signaling via estrogen receptor-alpha.* Nat Chem Biol, 2013. **9**(5): p. 326-32.

50. Motlagh, H.N. and V.J. Hilser, *Agonism/antagonism switching in allosteric ensembles.* Proc Natl Acad Sci U S A, 2012. **109**(11): p. 4134-9.

51. Merski, M., et al., *Homologous ligands accommodated by discrete conformations of a buried cavity.* Proc Natl Acad Sci U S A, 2015. **112**(16): p. 5039-44.

52. Warne, T., et al., *The structural basis for agonist and partial agonist action on a beta(1)-adrenergic receptor.* Nature, 2011. **469**(7329): p. 241-4.

53. Dalvit, C., *NMR methods in fragment screening: theory and a comparison with other biophysical techniques.* Drug Discov Today, 2009. **14**(21-22): p. 1051-7.

54. Jordan, J.B., et al., *Fragment based drug discovery: practical implementation based on (1)(9)F NMR spectroscopy.* J Med Chem, 2012. **55**(2): p. 678-87.

55. Fielding, L., *NMR methods for the determination of protein-ligand dissociation constants.* Curr Top Med Chem, 2003. **3**(1): p. 39-53.

56. Dalvit, C., et al., *Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability.* J Am Chem Soc, 2003. **125**(25): p. 7696-703.

57. Dai, H., et al., *SIRT1 activation by small molecules: kinetic and biophysical evidence for direct interaction of enzyme and activator.* J Biol Chem, 2010. **285**(43): p. 32695-703.

58. Gloster, T.M., et al., *Structural studies of the beta-glycosidase from Sulfolobus solfataricus in complex with covalently and noncovalently bound inhibitors.* Biochemistry, 2004. **43**(20): p. 6101-9.

59. Johnson, E.A., et al., *Agonist-selective mechanisms of mu-opioid receptor desensitization in human embryonic kidney 293 cells.* Mol Pharmacol, 2006. **70**(2): p. 676-85.

60. Urban, J.D., et al., *Functional selectivity and classical concepts of quantitative pharmacology.* J Pharmacol Exp Ther, 2007. **320**(1): p. 1-13.

61. Raehal, K.M., et al., *Functional selectivity at the mu-opioid receptor: implications for understanding opioid analgesia and tolerance.* Pharmacol Rev, 2011. **63**(4): p. 1001-19.

62. Pradhan, A.A., et al., *Agonist-Specific Recruitment of Arrestin Isoforms Differentially Modify Delta Opioid Receptor Function.* J Neurosci, 2016. **36**(12): p. 3541-51.

63. Reinhart, G.D., *The Determination of Thermodynamic Allosteric Parameters of an Enzyme Undergoing Steady-State Turnover.* Arch Biochem Biophys, 1983. **224**(1): p. 389-401.

64. Maehle, A.H., C.R. Prull, and R.F. Halliwell, *The emergence of the drug receptor theory.* Nat Rev Drug Discov, 2002. **1**(8): p. 637-41.

65. Kirk, K.L., *Selective fluorination in drug design and development: an overview of biochemical rationales.* Curr Top Med Chem, 2006. **6**(14): p. 1447-56.

66. Dimagno, S.G. and H. Sun, *The strength of weak interactions: aromatic fluorine in drug design.* Curr Top Med Chem, 2006. **6**(14): p. 1473-82.

67. Kuntz, I.D., et al., *The maximal affinity of ligands.* Proc Natl Acad Sci U S A, 1999. **96**(18): p. 9997-10002.

68. Tullman, J., et al., *Enzymatic protein switches built from paralogous input domains.* Biotechnol Bioeng, 2016. **113**(4): p. 852-8.

69. Wacker, D., et al., *Structural Features for Functional Selectivity at Serotonin Receptors.* Science, 2013. **340**(6132): p. 615-619.

70. Chang, S.D. and M.R. Bruchas, *Functional Selectivity at GPCRs: New Opportunities in Psychiatric Drug Discovery.* Neuropsychopharmacology, 2014. **39**(1): p. 248-249.

71. Pereira, D.A. and J.A. Williams, *Origin and evolution of high throughput screening.* Br J Pharmacol, 2007. **152**(1): p. 53-61.

72. Macarron, R., et al., *Impact of high-throughput screening in biomedical research.* Nat Rev Drug Discov, 2011. **10**(3): p. 188-95.

73. Hajduk, P.J., et al., *Privileged molecules for protein binding identified from NMR-based screening.* J Med Chem, 2000. **43**(18): p. 3443-7.

74. Hajduk, P.J. and J. Greer, *A decade of fragment-based drug design: strategic advances and lessons learned.* Nat Rev Drug Discov, 2007. **6**(3): p. 211-9.

75. Murray, C.W. and D.C. Rees, *The rise of fragment-based drug discovery.* Nat Chem, 2009. **1**(3): p. 187-92.

76. Yu, W. and A.D. MacKerell, Jr., *Computer-Aided Drug Design Methods.* Methods Mol Biol, 2017. **1520**: p. 85-106.

77. Wells, J.A. and C.L. McClendon, *Reaching for high-hanging fruit in drug discovery at protein-protein interfaces.* Nature, 2007. **450**(7172): p. 1001-9.

78. Ashkenazi, A., et al., *From basic apoptosis discoveries to advanced selective BCL-2 family inhibitors.* Nat Rev Drug Discov, 2017. **16**(4): p. 273-284.

79. Shuker, S.B., et al., *Discovering high-affinity ligands for proteins: SAR by NMR.* Science, 1996. **274**(5292): p. 1531-4.

80. Mattos, C., et al., *Multiple solvent crystal structures: probing binding sites, plasticity and hydration.* J Mol Biol, 2006. **357**(5): p. 1471-82.

81. Erlanson, D.A., et al., *Site-directed ligand discovery.* Proc Natl Acad Sci U S A, 2000. **97**(17): p. 9367-72.

82. Erlanson, D.A., J.A. Wells, and A.C. Braisted, *Tethering: fragment-based drug discovery.* Annu Rev Biophys Biomol Struct, 2004. **33**: p. 199-223.

83. Lodge, J.M., et al., *FP Tethering: a screening technique to rapidly identify compounds that disrupt protein-protein interactions.* Medchemcomm, 2014. **5**: p. 370-375.

84. Graham, S.E., N. Leja, and H.A. Carlson, *MixMD Probeview: Robust Binding Site Prediction from Cosolvent Simulations.* J Chem Inf Model, 2018. **58**(7): p. 1426-1433.

85. Faller, C.E., et al., *Site Identification by Ligand Competitive Saturation (SILCS) simulations for fragment-based drug design.* Methods Mol Biol, 2015. **1289**: p. 75-87.

86. Guvench, O. and A.D. MacKerell, Jr., *Computational fragment-based binding site identification by ligand competitive saturation.* PLoS Comput Biol, 2009. **5**(7): p. e1000435.

87. Yu, W., et al., *Pharmacophore modeling using site-identification by ligand competitive saturation (SILCS) with multiple probe molecules.* J Chem Inf Model, 2015. **55**(2): p. 407-20.

88. Bowman, G.R., et al., *Discovery of multiple hidden allosteric sites by combining Markov state models and experiments.* Proc Natl Acad Sci U S A, 2015. **112**(9): p. 2734-9.

89. Hart, K.M., et al., *Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators.* PLoS One, 2017. **12**(6): p. e0178678.

90. Guo, Z., et al., *Target the More Druggable Protein States in a Highly Dynamic Protein--Protein Interaction System.* J Chem Inf Model, 2016. **56**(1): p. 35-45.

91. Kalenkiewicz, A., B.J. Grant, and C.Y. Yang, *Enrichment of druggable conformations from apo protein structures using cosolvent-accelerated molecular dynamics.* Biology (Basel), 2015. **4**(2): p. 344-66.

92. Comitani, F. and F.L. Gervasio, *Exploring Cryptic Pockets Formation in Targets of Pharmaceutical Interest with SWISH.* J Chem Theory Comput, 2018. **14**(6): p. 3321-3331.

93. Johnson, D.K. and J. Karanicolas, *Ultra-High-Throughput Structure-Based Virtual Screening for Small-Molecule Inhibitors of Protein-Protein Interactions.* J Chem Inf Model, 2016. **56**(2): p. 399-411.

94. Johnson, D.K. and J. Karanicolas, *Selectivity by small-molecule inhibitors of protein interactions can be driven by protein surface fluctuations.* PLoS Comput Biol, 2015. **11**(2): p. e1004081.

95. Bohnuud, T., D. Kozakov, and S. Vajda, *Evidence of conformational selection driving the formation of ligand binding sites in protein-protein interfaces.* PLoS Comput Biol, 2014. **10**(10): p. e1003872.

96. Leaver-Fay, A., et al., *ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules.* Methods Enzymol, 2011. **487**: p. 545-74.

97. Rubenstein, A.B., et al., *Systematic Comparison of Amber and Rosetta Energy Functions for Protein Structure Evaluation.* J Chem Theory Comput, 2018. **14**(11): p. 6015-6025.

98. Hawkins, P.C., A.G. Skillman, and A. Nicholls, *Comparison of shape-matching and docking as virtual screening tools.* J Med Chem, 2007. **50**(1): p. 74-82.

99. Arkin, M.R., Y. Tang, and J.A. Wells, *Small-molecule inhibitors of protein-protein interactions: progressing toward the reality.* Chem Biol, 2014. **21**(9): p. 1102-14.

100. Scott, D.E., et al., *Small molecules, big targets: drug discovery faces the protein-protein interaction challenge.* Nat Rev Drug Discov, 2016. **15**(8): p. 533-50.

101. Jubb, H., T.L. Blundell, and D.B. Ascher, *Flexibility and small pockets at protein-protein interfaces: New insights into druggability.* Prog Biophys Mol Biol, 2015. **119**(1): p. 2-9.

102. Acoca, S., et al., *Molecular dynamics study of small molecule inhibitors of the Bcl-2 family.* Proteins, 2011. **79**(9): p. 2624-36.

103. Souers, A.J., et al., *ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets.* Nat Med, 2013. **19**(2): p. 202-8.

104. Petros, A.M., et al., *Fragment-based discovery of potent inhibitors of the anti-apoptotic MCL-1 protein.* Bioorg Med Chem Lett, 2014. **24**(6): p. 1484-8.

105. Kussie, P.H., et al., *Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain.* Science, 1996. **274**(5289): p. 948-53.

106. Zhao, Y., et al., *Small-molecule inhibitors of the MDM2-p53 protein-protein interaction (MDM2 Inhibitors) in clinical trials for cancer treatment.* J Med Chem, 2015. **58**(3): p. 1038-52.

107. Arkin, M.R., et al., *Binding of small molecules to an adaptive protein-protein interface.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1603-8.

108. Bordner, A.J., *Predicting small ligand binding sites in proteins using backbone structure.* Bioinformatics, 2008. **24**(24): p. 2865-71.

109. Guo, Z., et al., *Identification of protein-ligand binding sites by the level-set variational implicit-solvent approach.* J Chem Theory Comput, 2015. **11**(2): p. 753-65.

110. Borrel, A., et al., *PockDrug: A Model for Predicting Pocket Druggability That Overcomes Pocket Estimation Uncertainties.* J Chem Inf Model, 2015. **55**(4): p. 882-95.

111. Hussein, H.A., et al., *PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins.* Nucleic Acids Res, 2015. **43**(W1): p. W436-42.

112. Cimermancic, P., et al., *CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites.* J Mol Biol, 2016. **428**(4): p. 709-719.

113. Afonine, P.V., et al., *FEM: feature-enhanced map.* Acta Crystallogr D Biol Crystallogr, 2015. **71**(Pt 3): p. 646-66.

114. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.

115. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* J Phys Chem B, 1998. **102**(18): p. 3586-616.

116. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics, 1983. **79**: p. 926-935.

117. Steinbach, P.J. and B.R. Brooks, *New spherical-cutoff methods for long-range forces in macromolecular simulation.* Journal of Computational Chemistry, 1994. **15**(7): p. 667-683.

118. Essman, U., et al., *A smooth particle mesh Ewald method.* Journal of Chemical Physics, 1995. **103**: p. 8577-8593.

119. Alexander, C. and E.T. Rietschel, *Invited review: Bacterial lipopolysaccharides and innate immunity.* Journal of Endotoxin Research, 2001. **7**(3): p. 167-202.

120. Freeman, J.T.C. and W.C. Wimley, *TMBB-DB: a transmembrane β-barrel proteome database.* Bioinformatics, 2012. **28**(19): p. 2425-2430.

121. Storek, K.M., et al., *Monoclonal antibody targeting the beta-barrel assembly machine of Escherichia coli is bactericidal.* Proc Natl Acad Sci U S A, 2018. **115**(14): p. 3692-3697.

122. Dankert, J., S.M. Hammond, and W.A. Cramer, *Reversal by trypsin of the inhibition of active transport by colicin E1.* J Bacteriol, 1980. **143**(2): p. 594-602.

123. Zakharov, S.D. and W.A. Cramer, *Colicin crystal structures: pathways and mechanisms for colicin insertion into membranes.* Biochim Biophys Acta, 2002. **1565**(2): p. 333-46.

124. Housden, N.G., et al., *Directed epitope delivery across the Escherichia coli outer membrane through the porin OmpF.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21412-7.

125. Yamashita, E., et al., *Crystal structures of the OmpF porin: function in a colicin translocon.* EMBO J, 2008. **27**(15): p. 2171-80.

126. Benedetti, H., et al., *Individual domains of colicins confer specificity in colicin uptake, in pore-properties and in immunity requirement.* Journal of Molecular Biology, 1991. **217**(3): p. 429-439.

127. Liu, A., et al., *Antibiotic Sensitivity Profiles Determined with an Escherichia coli Gene Knockout Collection: Generating an Antibiotic Bar Code.* Antimicrobial Agents and Chemotherapy, 2010. **54**(4): p. 1393-1403.

128. Swick, M.C., et al., *Expression of Multidrug Efflux Pump Genes acrAB-tolC, mdfA, and norE in Escherichia coli Clinical Isolates as a Function of Fluoroquinolone and Multidrug Resistance.* Antimicrobial Agents and Chemotherapy, 2011. **55**(2): p. 921-924.

129. Besingi, R.N. and P.L. Clark, *Extracellular protease digestion to evaluate membrane protein cell surface localization.* Nat Protoc, 2015. **10**(12): p. 2074-80.

130. Kleanthous, C., P. Rassam, and C.G. Baumann, *Protein-protein interactions and the spatiotemporal dynamics of bacterial outer membrane proteins.* Curr Opin Struct Biol, 2015. **35**: p. 109-15.

131. Chavent, M., et al., *How nanoscale protein interactions determine the mesoscale dynamic organisation of bacterial outer membrane proteins.* Nat Commun, 2018. **9**(1): p. 2846.

132. White, P., et al., *Exploitation of an iron transporter for bacterial protein antibiotic import.* Proc Natl Acad Sci U S A, 2017. **114**(45): p. 12051-12056.

133. Bohnert, J.A., B. Karamian, and H. Nikaido, *Optimized Nile Red efflux assay of AcrAB-TolC multidrug efflux system shows competition between substrates.* Antimicrob Agents Chemother, 2010. **54**(9): p. 3770-5.

134. Seeger, M.A., et al., *Engineered disulfide bonds support the functional rotation mechanism of multidrug efflux pump AcrB.* Nat Struct Mol Biol, 2008. **15**(2): p. 199-205.

135. Bohnert, J.A., et al., *Determination of real-time efflux phenotypes in Escherichia coli AcrB binding pocket phenylalanine mutants using a 1,2'-dinaphthylamine efflux assay.* PLoS One, 2011. **6**(6): p. e21196.

136. Iyer, R., et al., *A fluorescent microplate assay quantifies bacterial efflux and demonstrates two distinct compound binding sites in AcrB.* Antimicrob Agents Chemother, 2015. **59**(4): p. 2388-97.

137. Misra, R., et al., *Importance of Real-Time Assays To Distinguish Multidrug Efflux Pump-Inhibiting and Outer Membrane-Destabilizing Activities in Escherichia coli.* J Bacteriol, 2015. **197**(15): p. 2479-88.

138. Cramer, W.A., O. Sharma, and S.D. Zakharov, *On mechanisms of colicin import: the outer membrane quandary.* Biochem J, 2018. **475**(23): p. 3903-3915.

139. Tuson, H.H. and J.S. Biteen, *Unveiling the inner workings of live bacteria using super-resolution microscopy.* Anal Chem, 2015. **87**(1): p. 42-63.

140. Sterling, T. and J.J. Irwin, *ZINC 15--Ligand Discovery for Everyone.* J Chem Inf Model, 2015. **55**(11): p. 2324-37.

141. Kabsch, W., *Automatic indexing of rotation diffraction patterns.* Journal of Applied Crystallography, 1988. **21**(1): p. 67-72.

142. Evans, P.R., *An introduction to data reduction: space-group determination, scaling and intensity statistics.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 282-92.

143. McCoy, A.J., et al., *Phaser crystallographic software.* J. Appl. Cryst., 2007. **40**: p. 658-674.

144. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

145. Krissinel, E., *Enhanced fold recognition using efficient short fragment clustering.* Journal of Molecular Biochemistry, 2012. **1**(2): p. 76-85.

146. Evans, P., *Scaling and assessment of data quality.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 72-82.

147. Diederichs, K. and P.A. Karplus, *Improved R-factors for diffraction data analysis in macromolecular crystallography.* Nat Struct Biol, 1997. **4**(4): p. 269-75.

148. Weiss, M.S., *Global indicators of X-ray data quality.* Journal of Applied Crystallography, 2001. **34**: p. 130-135.

149. Karplus, P.A. and K. Diederichs, *Linking crystallographic model and data quality.* Science, 2012. **336**(6084): p. 1030-3.

150. Evans, P., *Biochemistry. Resolving some old problems in protein crystallography.* Science, 2012. **336**(6084): p. 986-7.