Designing allosteric control into enzymes by chemical rescue of structure

By

Katelyn Deckert

Submitted to the graduate degree program in Molecular Biosciences and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

_____

Chairperson, Dr. John Karanicolas

_____

Dr. Mark Richter

_____

Dr. Eric Deeds

Date defended: December 10, 2012

The Thesis Committee for Katelyn Deckert

certifies that this is the approved version of the following thesis:

Designing allosteric control into enzymes by chemical rescue of structure

_____

Chairperson, Dr. John Karanicolas

Date approved: December 10, 2012

Abstract

In recent years chemical biology has been used to engineer ligand-dependent activity into enzymes to control a variety of signaling pathways, such as modulating protein production, degradation, and localization. These successes have generally incorporated a naturally allosteric domain to an enzyme of interest. Here, we demonstrate a novel approach for designing *de novo* allosteric effector sites directly into an enzyme. Our approach diverges from traditional chemical rescue in that it does not rely on the disruption and restoration of active site chemistry as a means to control enzyme function, but rather on the disruption and restoration of *structure*. We present two examples, W33G in a β-glycosidase enzyme (β-gly) and W492G in a β-glucuronidase enzyme (β-gluc), in which we engineer indole-dependent activity into enzymes by removing a buried tryptophan sidechain that serves as a buttress for the active site architecture, which results in a loss of activity for both enzymes. In both cases we demonstrate that subsequent addition of exogenous indole restores catalytic function. We demonstrate through analysis of enzyme kinetics that the rescued β-gly W33G enzyme is fully functionally equivalent to that of the wild type enzyme. We establish the structural basis for inactivation and rescue by presenting the apo and indole-bound crystal structures of β-gly W33G. Finally, we use this rationally designed switch to control β-glycosidase activity in living cells with the use of exogenous indole. Chemical rescue of protein structure may represent a general approach for designing allosteric control into enzymes, and thus may serve as a starting point for building a variety of bioswitches and sensors.

Acknowledgements

We thank S. Jimmy Budiardjo and Luke Brunner for initial production and tests of β-gluc. We thank Dr. Scott Lovell for solving crystal structures of apo and holo β-gly W33G. We thank Ning Zheng, Chet Egan, Eric Deeds, Jacob Corn, Audrey Lamb, Scott Hefty, Aron Fenton, Tanja Kortemme, and Brian Kuhlman for valuable discussions. We thank Marco Moracci for providing the β-gly gene, Bret Wallace and Matt Redinbo for providing the β-gluc gene, Susan Egan for providing the pHG165 vector, the COBRE PPG for assistance in subcloning, and Kevin Battaile for synchrotron data collection. 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 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 Center for Research Resources (5P20RR017708-07S1 and 5P20RR017708-09), the National Institute of General Medical Sciences (8 P20 GM103420-10), the NIH Dynamic Aspects of Chemical Biology Predoctoral Training Grant 2T32GM008545-17, and the Alfred P. Sloan Fellowship.

Table of contents

Introduction

The use of small molecules to control the activity of re-engineered enzymes is a powerful approach that can be used to control cell shape, signal transduction, growth, and survival [1-4]. The design strategy for introducing pharmacological control into enzymes underlying each of these successful examples has been based on the modularity of protein domains [5]. By fusing a naturally allosteric protein domain to a separate catalytic domain, ligand binding to the effector site can affect enzyme activity in select cases [1-4,6]. However, screening of multiple domain arrangements has been required in these examples to identify the new means of allosteric regulation of enzyme activity, which underscores the infeasibility of predicting the detailed mechanism of how effector binding results in altered activity [1-4,6]. In the case of a naturally allosteric single-domain enzyme, elucidating the structural basis for regulation can enable introduction of cysteine residues such that activity is dependent on redox conditions that induce disulfide bond formation [7]. A more general approach involves engineering new catalytic activity into a naturally allosteric (non-catalytic) domain [8], but challenges associated with current methods of computational enzyme design [9] limit the range of reactions which can be designed in this manner, resulting in only a single successful example to date.

Here we describe a complementary strategy to these approaches. Rather than utilizing a natural allosteric transition, we have instead developed a technique for engineering *de novo* allosteric control into a protein domain that is not naturally allosteric. Our approach is based on chemical rescue [10], in which an amino acid sidechain known to be critical for catalysis – often histidine – is deleted, leading to loss of enzymatic activity and the creation of an effector binding

site. Upon addition of a complementary exogenous ligand – such as imidazole – activity can be restored when the small molecule binds to the complementary effector site. Inspired by the demonstration that internal protein cavities resulting from mutations can accommodate small hydrophobic ligands [11,12], here we propose an alternate approach to chemical rescue. Rather than remove a surface sidechain required for *catalysis*, we instead remove a buried element required for *structure*. Our approach consists of removing a buried structural element in the enzyme, which in turn distorts the active site geometry and results in loss of function; subsequent exogenous replacement of the complementary small molecule can then be used to restore structure, and hence activity (**Figure 1a**). The ability of an effector ligand to restore protein activity will require not simply binding, but also precise rescue of structure. Previous attempts to rescue cavity-forming mutants with activating ligands have included a zinc finger transcription factor [13] and a hormone-receptor pair [14]. Each of these studies identified a rescuing ligand by screening restricted compound libraries, and demonstrated that the selected ligand restored partial activity. In neither case was activity fully restored, however; we propose that complete rescue requires a ligand with exquisite structural complementarity unlikely to be found in a screening library of limited size. We expect that complete rescue of function will require exact ligand-cavity complementarity, which we attain by rationally matching the effector ligand to the structure of the deleted moiety.

Though this approach may in principle be applied to a deleted structural element corresponding to sidechains from multiple amino acid residues, in this study we focus on mutation of a single buried tryptophan residue to glycine (W→G). If this tryptophan sidechain is serving as a structural "buttress" to prevent collapse of the enzyme active site, its removal will lead to distortion of the catalytic geometry and thus loss of enzyme activity. Replacement of the

buttress via addition of exogenous indole – which perfectly complements the deleted sidechain – will restore the original protein conformation, and thus rescue catalytic activity. Critically, the mutation site need not be at the active site. If the structural disruption caused by the remote mutation is transduced to the active site, the recovery of structure upon complementation is also expected to be transduced to the active site through allosteric activation. This approach therefore represents a rational strategy for designing remote *de novo* allosteric effector binding sites into proteins.

Materials and Methods

Protein expression and purification

Full-length (untagged) recombinant wild type *Sulfolobus solfataricus* β-glycosidase (β-gly) was expressed from a pET29 vector in *E. coli* Rosetta(DE3)pLysS cells at 37°C for 4 hours. The cells were resuspended in Lysis Buffer (50 mM sodium phosphate pH 6.8) and lysed with three freeze-thaw cycles. The cell lysate was centrifuged, and the supernatant was incubated at 75°C for 30 min. The aggregated cellular proteins were cleared by centrifugation at 27,000g for 30 min, leaving the β-gly in the supernatant, which was then further purified by HPLC using a HiLoad 16/60 Superdex 75 gel filtration column (GE Healthcare). The purified enzyme was pooled and dialyzed against β-gly Enzyme Assay Buffer (50 mM Sodium Phosphate pH 6.5).

Full length hexahistidine-tagged recombinant wild type *Escherichia coli* β-glucuronidase (β-gluc) was expressed from a pET28a vector in *E. coli* Rosetta 2(DE3)pLysS cells at 15ºC overnight. The cells were resuspended in Lysis Buffer (50 mM Tris, 150 mM NaCl, 5 mM imidazole pH 8.0) and sonicated for two minutes (Fisher Scientific Sonic Dismembrator Model

100). The cell lysates were then centrifuged at 22,000g for 30 min. The β-gluc remained in the supernatant, which was purified by HPLC affinity chromatography with Ni-chelated Sepharose Fast Flow Resin (GE Healthcare), followed by a HiLoad 16/60 Superdex 75 gel filtration column (GE Healthcare). The purified enzyme was pooled and dialyzed against β-gluc Enzyme Assay Buffer (50 mM sodium phosphate pH 7.4, 100 mM NaCl). All protein concentrations were determined with reference to bovine albumin standards using Bradford assays. Point mutations were introduced using the QuikChange methodology (Stratagene), and mutant proteins were purified as described for the corresponding wild-type protein.

Assays of enzyme kinetics

β-glycosidase assays were carried out at 37°C in 100 μL reactions in black 96-well plates. Each reaction contained 58 nM β-gly in Enzyme Assay Buffer (50 mM sodium phosphate pH 6.5), and indole concentrations ranging from 0 to 10 mM in 5% DMSO. Duplicate reactions were initiated by addition of 20 μL substrate fluorescein di-β-D-galactopyranoside (FDG) dissolved in 0.1% ethanol, 0.1% DMSO, and 99.8% water. FDG concentrations ranged from 1 to 750 μM for complete kinetic analysis (Figure 2), but were held fixed at 100 μM for initial evaluation of rescue (Figure 1). β-glycosidase cleaves FDG twice to yield one molecule of D-galactose and two molecules of fluorescein. Product formation (fluorescein) was continuously monitored for 15 min by fluorescence with an excitation filter of 485nm and an emission filter of 528nm using a Synergy 2 Microplate reader (Biotek). Initial velocities were calculated by linear regression using Microsoft Excel, and apparent Michaelis-Menten parameters were determined by direct fitting of the data to the Michaelis-Menten equation using KaleidaGraph. Fitting the allosteric rate equation to obtain the enzyme-indole dissociation constant was carried out with

C++ code invoking the GSL multidimensional minimization tools (http://www.gnu.org/s/gsl/). Source code to carry out this fitting will be freely provided upon request.

β-glucuronidase assays were carried out at 30ºC in 100 µL reactions in clear 96-well plates. Each reaction contained final concentrations of 40 mM 2-nitrophenyl-β-D-glucopyranoside (ONPGlc), 5% ethanol, 3.7 µM β-gluc in Enzyme Assay Buffer (75 mM sodium phosphate pH 7.4, 100 mM NaCl), and indole concentrations ranging from 0 to 10 mM. β-glucuronidase cleaves ONPGlc to yield D-glucose and 2-nitrophenol. Product formation (*ortho*-nitrophenol, ONP) was monitored continuously for 30 min by absorbance at 405 nm in a Synergy 2 Microplate reader (Biotek).

Qualitative (X-gal) assay of β-gly rescue in E. coli

E. coli XL1-Blue cells (Stratagene) were transformed with plasmids encoding either wild-type, W33G, or W425G β-gly under control of a *lac* promoter in the pHG165 vector. Cells were grown in 10 mL cultures of LB media supplemented with 100 µg/mL ampicillin at 37°C with vigorous shaking to an OD_{600} of 0.6. Protein expression was induced with 1 mM IPTG for 3 h. Each cell culture was split into two equal parts; one was supplemented with 2 mM indole and 0.5% DMSO while the other was supplemented only with 0.5% DMSO. From each culture, a 1 mL aliquot was removed and supplemented with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) to a final concentration of 196 µM. The remainder of the original cultures were reserved for further fractionation (**Figure S4**). All cultures were incubated an additional four hours, and photos were taken of the X-gal supplemented cultures.

The reserved cultures were then centrifuged to pellet the cells, and the growth media was removed. The cells were resuspended and washed in LB media three times, and lysed with three freeze-thaw cycles. The media and cell lysate fractions were further divided into equal aliquots, one of which was supplemented with indole to a final concentration of 2 mM. All media and cell lysate fractions were then supplemented with 196 μM X-gal and incubated for 4 h.

Quantitative (FDG) assay of β-gly rescue in E. coli

Cell cultures were grown as described above for the qualitative (X-gal) experiment. After 3 h of protein expression, cell density (OD_{600}) was normalized for both wild-type and W33G β-gly cultures. Each culture was divided into 1 mL aliquots and centrifuged to pellet the cells. The cells were resuspended in 1 mL of fresh LB media, then supplemented with either 1.5 mM indole and 0.5% DMSO or 0.5% DMSO with no indole. Four independent replicates of each culture were grown.

These cultures were then incubated for one hour at 37°C with vigorous shaking. For each 1 mL culture, half was removed and supplemented with 15 μM FDG in a volume of 50 μL (0.1% ethanol, 0.1% DMSO, and 99.8% water). The other half was supplemented with 50 μL of the equivalent buffer (0.1% ethanol, 0.1% DMSO, and 99.8% water) to serve as a blank for the fluorescence readings. All samples were incubated in the dark at 37°C with vigorous shaking for 1 h. From each sample, 100 μL were analyzed using a Synergy 2 Microplate reader (Biotek) in black 96-well plates. Fluorescein was detected with an excitation wavelength of 485nm and an emission wavelength of 528nm. Fluorescence of the used media in each culture was subtracted from the total fluorescence to give the amount of product (fluorescein) inside intact cells. The

fraction activity for β-gly W33G relative to wild-type was calculated by dividing the intracellular fluorescence between cells treated under analogous conditions.

Crystal structures of apo and holo β-gly W33G

Crystal structures for apo and holo β-gly W33G were solved in collaboration with Dr. Scott Lovell at the Protein Structure Laboratory at the University of Kansas.

β-gly W33G was concentrated to 10.0 mg/mL in HEPES pH 7.0 then 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 100 μL of the latter at 20 ºC. Prismatic crystals were obtained within 24-48 hours. To prepare the indole-bound complex, single crystals were transferred to a drop containing 100 mM indole dissolved in crystallization solution and incubated for 1 hour.

Single crystals of the apo form were transferred to a drop crystallization solution, which served as the cryoprotectant, before freezing in liquid nitrogen for data collection. Crystals soaked in the presence of indole were frozen directly from the soaking solution for data collection. Data were collected at the Advanced Photon Source IMCA-CAT beamline 17ID using a Dectris Pilatus 6M pixel array detector.

Intensities were integrated and scaled using the XDS [15] and Scala [16] packages respectively. Structure solution for unbound β-gly W33G was carried out by molecular replacement with Phaser [17] using the isomorphous non-crystallographic dimer of a previously

determined lower resolution structure of the W33G mutant (PDB: 3UUJ) as the search model. The higher resolution structures reported here were obtained during the revision stages for this manuscript. The earlier PDB entries 3UUJ and 3UUK were therefore withdrawn and replaced with the new structures (PDB: 4EAM and 4EAN) for the apo and indole bound structures respectively. Structure refinement and manual model building were performed with Phenix [18] and Coot [19] respectively. Anisotropic atomic displacement parameters were modeled in the final stages of refinement using a TLS model [20]. Structure validation was carried out using Molprobity [21].

In each structure, large difference density peaks (Fo-Fc) greater than 3σ were observed near Arg residues 286, 295 and 440 in each subunit and were originally modeled as water molecules. However, following refinement this site was covered with positive electron density (Fo-Fc) at a 5σ contour level indicating an underestimation of electrons. Therefore, Cl^- ions were modeled at these sites and no residual difference density was observed following refinement. Electron density consistent with 2-methyl-2,4-pentanediol (MPD) molecules was observed at various sites and modeled accordingly. Two Tris buffer molecules, likely acquired during purification, were also modeled to the Fo-Fc maps.

In the indole-bound structure, a large difference density (Fo-Fc) greater than 3σ and consistent with indole was observed in the active site and was modeled accordingly. Since the orientation of the indole could not be determined conclusively, it was modeled in the same orientation as the indole ring of W33 in the wild-type β-gly structure (PDB: 2CEQ).

In both structures disordered side chains were truncated to the point where difference electron density could be observed. Relevant crystallographic data are provided in **Table S1**.

Coordinates and structure factors for the apo and indole-bound crystal structures of β-gly W33G have been deposited with the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) with accession codes 4EAM and 4EAN, respectively.

Results

To test the hypothesis that protein function can be modulated through structure-based chemical rescue, we selected two divergent enzymes that cleave β-glycosidic bonds because of the ease associated with assaying the activity. While sequence-based evolutionary methods have proven useful in detecting energetic coupling [6,22], we instead relied on protein structure to identify buried tryptophan sidechains that may be serving to buttress the active site. Using available crystal structures [23-25], we selected Trp33 from the *S. solfataricus* β-glycosidase (β-gly) and Trp492 from *E. coli* β-glucuronidase (β-gluc) for testing this hypothesis (**Figure 1b,c**). While neither sidechain makes direct interactions with the substrate, each occupies a buried environment within 15 Å of the active site. Further, we find using modern packing metrics [26] that these two sidechains are among the 10% most tightly-packed buried tryptophan sidechains in the Protein Data Bank [27] suggesting their surroundings may be well suited to serve as indole binding sites. In essence, selective pressure on the amino acids surrounding each of these tryptophan sidechains has inadvertently evolved an optimal indole binding site given the constraints of the protein architecture required for function.

We found that both β-gly W33G and β-gluc W492G exhibit strongly diminished β-glycosidase activity relative to their respective wild-type enzymes. Remarkably, in both cases addition of indole led to dose-dependent recovery of enzyme activity (**Figure 1d,e**). In contrast,

we found that addition of indole up to 15 mM, the solubility limit of indole in aqueous solution, did not restore activity to four separate W→G point mutants of β-gly (W151G, W361G, W425G, W433G) or to a W→G point mutant of β-gluc (W471G) if the mutation was located at the active site. In the presence of 5 mM indole, β-gly W33G led to product formation at a rate 0.58 times that of the corresponding wild-type enzyme without indole (0.34 μM/min vs. 0.58 μM/min, **Figure 1d**); in the presence of 5 mM indole β-gluc W492G led to product formation at a rate 0.14 times that of the corresponding wild-type enzyme without indole (7.8 μM/min vs. 56 μM/min, **Figure 1e**). Because rescue of β-gly W33G was more effective than β-gluc W492G at this indole concentration, we selected the former for further characterization. While the studies below serve to demonstrate that the mechanism of inactivation and rescue for β-gly W33G are as designed (**Figure 1a**), we note that in the absence of additional experiments we cannot definitively establish the same of β-gluc W492G.

We first explored the mechanism of rescue by examining the rate of product formation as a function of both indole and substrate concentrations. In the simplest allosteric kinetic mechanism (**Figure S2**), the rate equation predicts that initial velocities as a function of substrate concentration can be fit to a rectangular hyperbola if the concentration of the allosteric modulator is fixed [28]. Indeed we find this to hold across a wide range of indole concentrations, allowing us to fit "apparent" values of the steady-state Michaelis-Menten kinetic parameters. Individually fitting experimental data collected at each indole concentration, we observe a sharp decrease in apparent K_m, from 1080 ± 30 μM without indole to 53 ± 3 μM at 10 mM indole, accompanied by an increase in apparent k_{cat}, from 0.00675 ± 0.00004 s^{-1} without indole to 0.234 ± 0.009 s^{-1} at 10 mM indole (**Figure 2a,b**). By contrast, the kinetic parameters of the wild-type enzyme are nearly indole-independent, with a K_m of 58 ± 4 μM and k_{cat} of 0.265 ± 0.002 s^{-1} without indole, and a K_m

of 48 ± 1 μM and k_{cat} of 0.221 ± 0.006 s^{-1} at 10 mM indole, confirming that modulation of β-gly W33G activity indeed relies on the W→G point mutation. Notably, both K_m and k_{cat} of β-gly W33G reach the corresponding wild-type values at high indole concentration, suggesting that the rescued holo enzyme is fully functionally equivalent to the wild-type enzyme. Relative to the apo form of the engineered enzyme, addition of indole leads to a 20-fold decrease in K_m and a 39-fold increase in k_{cat}; these two ratios, termed Q and W respectively [28], define the allosteric linkage between substrate and allosteric activator. In other words, from the value of Q we infer that the presence of saturating indole enhances substrate binding by -1.8 kcal/mol, or equivalently that the presence of saturating substrate enhances indole binding by -1.8 kcal/mol. Collectively, the strong linkage through both Q and W together produce a 780-fold change in the ratio k_{cat} / K_m upon activation by saturating concentrations of indole.

To further test the appropriateness of this simple allosteric kinetic mechanism (**Figure S2**) for describing activation of β-gly W33G by indole, we fit the rate equation for this model [28] to the complete set of experimentally-determined initial velocities. The limiting values of the Michaelis-Menten kinetic parameters for the apo and holo enzyme were set to those of the W33G and wild-type enzyme in the absence of indole, leaving a single free parameter in the fitting corresponding to the enzyme-indole dissociation constant in the absence of substrate. This simple kinetic model proved sufficient to describe the initial velocity across a broad range of indole and substrate concentrations (**Figure 2c**). The value of the free parameter in this optimal fit was 15 mM, representing the enzyme-indole dissociation constant in the *absence* of substrate. From the linkage relationship, this dissociation constant in the *presence* of saturating substrate drops to 0.75 mM.

To further understand the detailed mechanism for β-gly W33G rescue, we next solved the crystal structure of β-gly W33G to 1.7 Å resolution (**Table S1**). The structural basis for inactivation of β-gly W33G is clearly demonstrated by comparing this crystal structure to that of the wild-type enzyme with a substrate analog bound in the active site (**Figure 3a**). In the β-gly W33G structure, the cavity resulting from mutation of W33 to glycine is occupied by W433, a nearby tryptophan, which moves away from its original position near the active site. Though no substrate analog is present in our structure, previous studies of wild-type β-gly have highlighted the importance of W433 in forming a key hydrogen bond to substrate hydroxyl group [24]. Consistent with our analysis of enzyme kinetics, studies of a closely related enzyme found that a synthetic substrate lacking this hydroxyl group exhibits a markedly higher K_m than the corresponding natural substrate [29].

To explore the basis for re-activation of β-gly W33G, we then soaked these crystals with indole and solved the crystal structure to 1.75 Å resolution (**Table S1**). Dramatically, in this structure we find that W433 has reverted to its original location near the active site (**Figure 3b**). As expected, we further observe clear electron density in the void left by the W33G mutation, which is presumably occupied by indole (**Figures 3c and S3**). An indole molecule modeled into this void fits the electron density map, which closely aligns with the W33 side chain in the wild-type structure. Aside from very small changes near the W33G mutation, the bound and unbound structures are otherwise identical (0.22 Å Cα RMSD excluding residue 33). The reversion of the indole-bound β-gly W33G structure to match that of the wild-type enzyme is unsurprising given that it is fully consistent with our kinetic characterization showing complete rescue of β-gly W33G at high indole concentration.

To determine whether our engineered allosteric enzyme is active in living cells, we next grew *E. coli* cells containing expression plasmids in media supplemented with one of two substrates: either X-gal for qualitative (visual) detection (**Figure 4a**), or FDG for quantitative (fluorometric) detection (**Figure 4b**). We found that cell cultures expressing the wild-type enzyme retained β-glycosidase activity in the presence or absence of indole. In contrast, β-gly W425G, an active-site mutation that had no detectable activity *in vitro* (**Figure S1**), served as a negative control with no detectable activity in cell culture with or without indole. Cell cultures expressing β-gly W33G without the addition of indole exhibited a significant reduction in activity relative to the wild-type enzyme. Notably, activity in cell cultures expressing β-gly W33G could be rescued by addition of exogenous indole to the media (**Figure 4a,b**). To exclude the possibility that indole was disrupting cells and thus releasing β-gly W33G into the media, we tested for β-glycosidase activity in the media both quantitatively using FDG and qualitatively using X-gal. By partitioning components of these cultures we tracked β-glycosidase activity and found that intracellular activation of β-gly W33G was responsible for product formation (**Figure S4**). By interpolation from indole dependence of β-gly W33G determined in the analogous *in vitro* experiment (**Figure 4c**), we find that addition of 1.5 mM indole to the media leads to intracellular β-gly W33G activation commensurate with approximately 1.35 mM indole *in vitro* (**Figure 4b**). The extent of β-gly W33G activation, corresponding to 27% relative to the wild-type enzyme, nearly reaches the maximal level expected at this concentration of exogenous indole.

Discussion

In both β-gly W33G and β-gluc W492G, the structural disruption arising from the cavity-forming mutations was transduced from a remote location to the active site, leading to diminished function of the engineered apo enzyme. Identifying analogous sites in other enzymes represents a key challenge in extending this approach to introduce allosteric control into other enzymes. Coevolving networks of amino acids have been used to trace energetic "wires" linking remote sites in proteins, which may allow rapid detection of mutations at distant regions that could be used to modulate protein function [6,30]. Alternatively at the level of structure, complementary predictions may be facilitated computationally to predict whether the structural disruption arising from a particular cavity-forming mutation will be transduced to the active site [31-35]. From a number of studies, the identification of allosteric sites with no known natural effector in a variety of protein families has fueled speculation that energetic coupling of functional sites to remote regions of the protein may be a common phenomenon [36]. The suggestion that proteins are "primed" to show allosteric behavior implies that the current challenge in engineering switchable enzymes lies not in designing allosteric mechanisms to transduce signals from a remote site, but rather in incorporating *de novo* binding sites [37]. This view is consistent with the results presented here: the effector binding site was rationally designed by chemical rescue of structure, allowing allosteric modulation of enzyme activity to naturally emerge.

The work presented here extends previous demonstrations of using small molecules that complement protein cavities as a way to control protein function [11,12]. We anticipate that switchable enzymes can be utilized to study the biophysical properties of proteins in response to

cavity mutations and the subsequent binding of small molecules. Our method measures protein function as the readout for binding of small molecules, as opposed to only detecting binding. The relative binding of various small molecules in a protein cavity can be detected and compared by how well they restore function, which correlates to how well they restore structure. This could be why our approach of rationally matching the effector to the deleted structural moiety works better than previous studies where they screen from a limited library of small molecules that are not perfectly complementary to the effector site [14].

Structural disruption by mutation of a single buried tryptophan residue to glycine naturally lends itself to the use of indole as its cognate activating effector ligand. However, indole is a bioactive molecule present in many cell types and may prove limiting for certain future applications. Further, the size and chemical characteristics of indole may place intrinsic limits on binding affinity, manifest through the high concentrations of indole required for rescue in the two examples presented here. However, chemical rescue of structure need not be limited to this particular mutation or even to cavity-forming mutations at a single site. We anticipate that by identifying constellations of atoms that match a particular compound, it is possible to create cavities in proteins that will be complemented by larger and more diverse compounds. Slightly larger and more hydrophobic ligands will lead to better binding affinity, which in turn may offer enhanced sensitivity and selectivity. As such, we expect that chemical rescue of structure will represent a technique for building molecular switches and sensors that respond to a wide variety of activating effector ligands, including those of biological interest. Using this as a starting point, it may be possible to use computational protein design or directed evolution to shift the effector ligand to another molecule, which in turn may allow development of a number of catalytically amplified biosensors.

Bibliography

(1) Karginov, A. V.; Ding, F.; Kota, P.; Dokholyan, N. V.; Hahn, K. M. *Nat Biotechnol* **2010**, *28*, 743.

(2) Yeh, B. J.; Rutigliano, R. J.; Deb, A.; Bar-Sagi, D.; Lim, W. A. *Nature* **2007**, *447*, 596.

(3) Tucker, C. L.; Fields, S. *Nat Biotechnol* **2001**, *19*, 1042.

(4) Guntas, G.; Mansell, T. J.; Kim, J. R.; Ostermeier, M. *Proc Natl Acad Sci U S A* **2005**, *102*, 11224.

(5) Ostermeier, M. *Curr Opin Struct Biol* **2009**, *19*, 442.

(6) Lee, J.; Natarajan, M.; Nashine, V. C.; Socolich, M.; Vo, T.; Russ, W. P.; Benkovic, S. J.; Ranganathan, R. *Science* **2008**, *322*, 438.

(7) Nomura, A. M.; Marnett, A. B.; Shimba, N.; Dotsch, V.; Craik, C. S. *Nat Struct Mol Biol* **2005**, *12*, 1019.

(8) Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; Degrado, W. F. *Proc Natl Acad Sci U S A* **2011**, *108*, 6823.

(9) Baker, D. *Protein Sci* **2010**, *19*, 1817.

(10) Toney, M. D.; Kirsch, J. F. *Science* **1989**, *243*, 1485.

(11) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. *Nature* **1992**, *355*, 371.

(12) Das, A.; Wei, Y.; Pelczer, I.; Hecht, M. H. *Protein Sci* **2011**, *20*, 702.

(13) Lin, Q.; Barbas, C. F., 3rd; Schultz, P. G. *J Amer Chem Soc* **2003**, *125*, 612.

(14) Guo, Z.; Zhou, D.; Schultz, P. G. *Science* **2000**, *288*, 2042.

(15) Kabsch, W. *J Appl Cryst* **1988**, *21*, 67.

(16) Evans, P. R. *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 282.

(17) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J Appl Crystallogr* **2007**, *40*, 658.

(18) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213.

(19) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486.

(20) Painter, J.; Merritt, E. A. *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 439.

(21) Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 12.

(22) Lockless, S. W.; Ranganathan, R. *Science* **1999**, *286*, 295.

(23) Gloster, T. M.; Roberts, S.; Perugino, G.; Rossi, M.; Moracci, M.; Panday, N.; Terinek, M.; Vasella, A.; Davies, G. J. *Biochemistry* **2006**, *45*, 11879.

(24) Gloster, T. M.; Roberts, S.; Ducros, V. M.; Perugino, G.; Rossi, M.; Hoos, R.; Moracci, M.; Vasella, A.; Davies, G. J. *Biochemistry* **2004**, *43*, 6101.

(25) Wallace, B. D.; Wang, H.; Lane, K. T.; Scott, J. E.; Orans, J.; Koo, J. S.; Venkatesh, M.; Jobin, C.; Yeh, L. A.; Mani, S.; Redinbo, M. R. *Science* **2010**, *330*, 831.

(26) Sheffler, W.; Baker, D. *Protein Sci* **2009**, *18*, 229.

(27) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J Mol Biol* **1977**, *112*, 535.

(28) Reinhart, G. D. *Arch Biochem Biophys* **1983**, *224*, 389.

(29) Namchuk, M. N.; Withers, S. G. *Biochemistry* **1995**, *34*, 16194.

(30) Reynolds, K. A.; McLaughlin, R. N.; Ranganathan, R. *Cell* **2011**, *147*, 1564.

(31) Machicado, C.; Bueno, M.; Sancho, J. *Protein Eng* **2002**, *15*, 669.

(32) Demerdash, O. N.; Daily, M. D.; Mitchell, J. C. *PLoS Comput Biol* **2009**, *5*, e1000531.

(33) Dixit, A.; Verkhivker, G. M. *PLoS Comput Biol* **2011**, *7*, e1002179.

(34) Kidd, B. A.; Baker, D.; Thomas, W. E. *PLoS Comput Biol* **2009**, *5*, e1000484.

(35) Laine, E.; Goncalves, C.; Karst, J. C.; Lesnard, A.; Rault, S.; Tang, W. J.; Malliavin, T. E.; Ladant, D.; Blondel, A. *Proc Natl Acad Sci U S A* **2010**, *107*, 11277.

(36) Hardy, J. A.; Wells, J. A. *Curr Opin Struct Biol* **2004**, *14*, 706.

(37) Wright, C. M.; Heins, R. A.; Ostermeier, M. *Curr Opin Chem Biol* **2007**, *11*, 342.

(38) Pinero-Fernandez, S.; Chimerel, C.; Keyser, U. F.; Summers, D. K. *J Bacteriol* **2011**, *193*, 1793.

(39) Lee, J. H.; Lee, J. *FEMS Microbiol Rev* **2010**, *34*, 426.

(40) Diederichs, K.; Karplus, P. A. *Nat Struct Biol* **1997**, *4*, 269.

(41) Weiss, M. S. *J Appl Cryst* **2001**, *34*, 130.

Figure Captions

Figure 1: Indole rescue. (A) Mutation of a remote "buttressing" tryptophan leads to structural disruption of catalytic geometry, resulting in loss of activity. Subsequent addition of exogenous indole restores structure, and thus activity. **(B)** Crystal structure of β-gly with a substrate analog (2-deoxy-2-fluoro-glucose, purple spheres) [24], showing the "buttressing" tryptophan (Trp33, orange) and a nearby tryptophan (Trp433, yellow). **(C)** Crystal structure of β-gluc [25] with a substrate analog (glucaro-δ-lactam, blue spheres) showing the "buttressing" tryptophan (Trp492, orange). **(D)** Indole-dependent activity of β-gly W33G, measured spectrofluorometrically using 100 μM fluorescein di-β-D-galactopyranoside (FDG) as a substrate. **(E)** Indole-dependent activity of β-gluc W492G, measured spectrophotometrically using 40 mM 2-nitrophenyl-β-D-glucopyranoside (ONPGlc) as a substrate. No indole was included in data for either wild-type enzyme shown here.

Figure 2: Rescued β-gly W33G is functionally equivalent to the wild-type enzyme. The apparent Michaelis-Menten parameters **(A)** K_m and **(B)** k_{cat} are presented as a function of indole for the wild type enzyme and W33G β-gly, measured spectrofluorometrically using FDG as a substrate. Parameters for wild type β-gly are essentially indole independent, and the β-gly W33G parameters approach the equivalent wild type values with increasing indole. Wild type and W33G data points overlap at 10 mM indole. Measurements were carried out in duplicate; error bars represent the standard error of the mean. **(C)** The measured initial velocities are consistent

with a simple allosteric model with one free parameter corresponding to the enzyme-indole dissociation constant.

Figure 3: Structural basis for indole rescue of β-gly W33G. Crystal structures of apo β-gly W33G and indole-bound β-gly W33G were refined to 1.7 Å resolution and 1.75 Å resolution, respectively. Structures are superposed to the wild type enzyme (pale green) [24]. A substrate analog (2-deoxy-2-fluoro-glucose, purple spheres) present in the structure of the wild type enzyme is shown to indicate the location of the active site. **(A)** Crystal structure of apo β-gly W33G (pink). Trp433 occupies the cavity produced by mutation of Trp33 to Gly, causing it to lose contact with the substrate. **(B)** Crystal structure of holo β-gly W33G (pink). Trp433 reverts to the conformation observed in the wild type enzyme. No other sidechains differ between the holo and apo structures of β-gly W33G. **(C)** A difference electron density (Fo-Fc) map contoured at 3σ from holo β-gly W33G chain A before inserting the indole. The electron density is consistent with indole (gray) at a location overlapping Trp33 in the wild type enzyme (pale green). The orientation of indole cannot be unambiguously determined at this resolution.

Figure 4: Indole rescue in living cells. (A) *E. coli* cells expressing β-gly supplemented with X-gal substrate for qualitative detection of activity. Addition of indole (indicated with +/-) affects neither wild type β-gly activity (left pair, positive control) nor β-gly W425G activity (right pair, negative control). Only β-gly W33G (center) shows indole-dependent activity. All tubes contain 196 μM X-gal and 1 mM IPTG; those marked with a "+" additionally contain 2 mM indole. **(B)**

Quantification of β-gly activity in *E. coli* cells supplemented with FDG substrate. Fraction of product formation by β-gly W33G after 30 min is reported relative to wild type β-gly. Four experimental replicates were averaged; error bars represent the standard error. **(C)** The analogous experiment was carried out *in vitro* to determine the maximal activity relative to the wild type enzyme that could be expected from addition of 1.5 mM indole to the cell cultures.

Figures

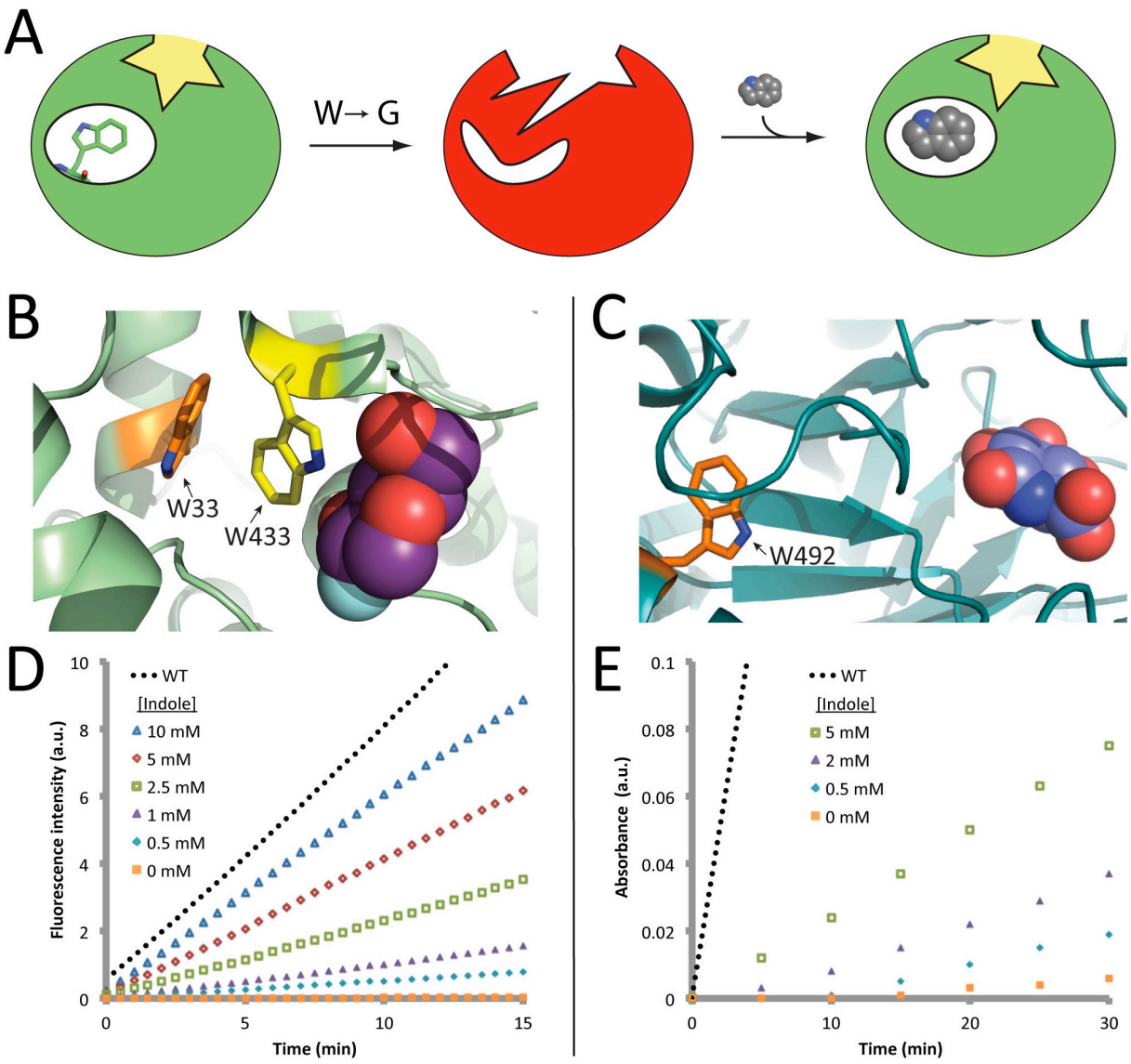

Figure 1

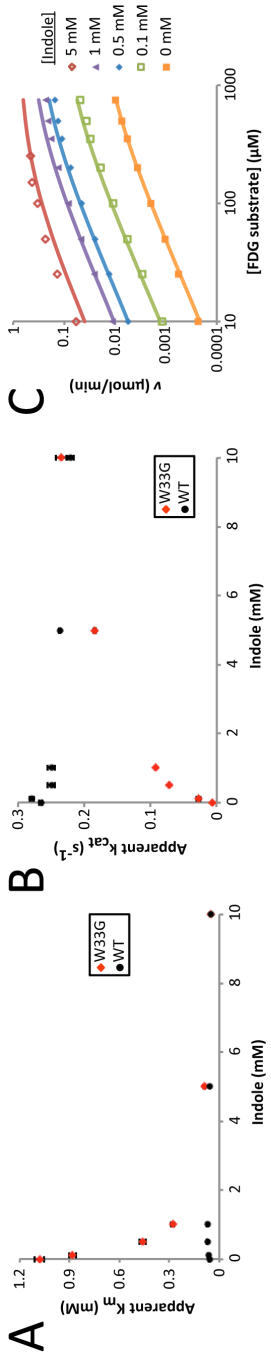
A
Apparent K_m (mM)
1.2
0.9
0.6
0.3
0
0
2
4
6
8
10
Indole (mM)
W33G
WT
B
Apparent k_{cat} (s^{-1})
0.3
0.2
0.1
0
0
2
4
6
8
10
Indole (mM)
W33G
WT
C
v (μmol/min)
1
0.1
0.01
0.001
0.0001
10
100
1000
[FDG substrate] (μM)
[Indole]
5 mM
1 mM
0.5 mM
0.1 mM
0 mM

Figure 2

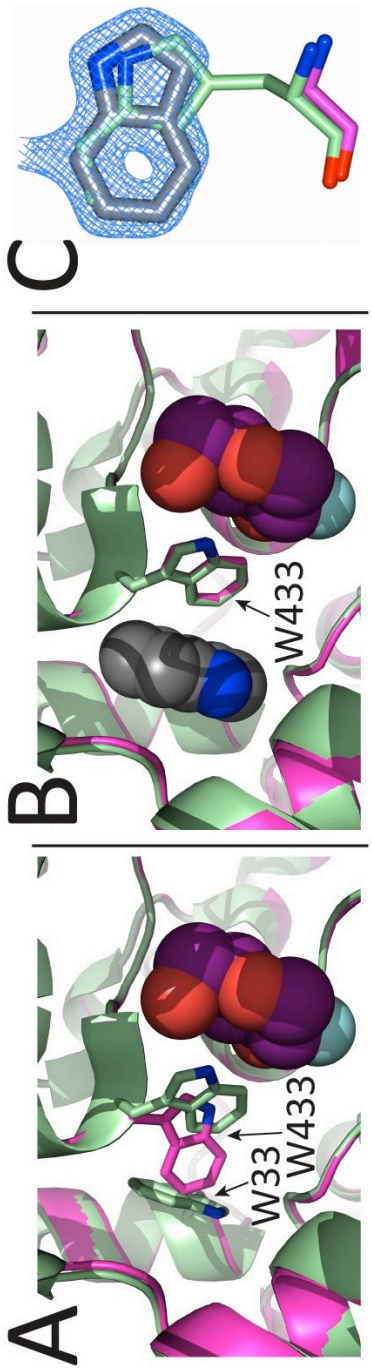

Figure 3

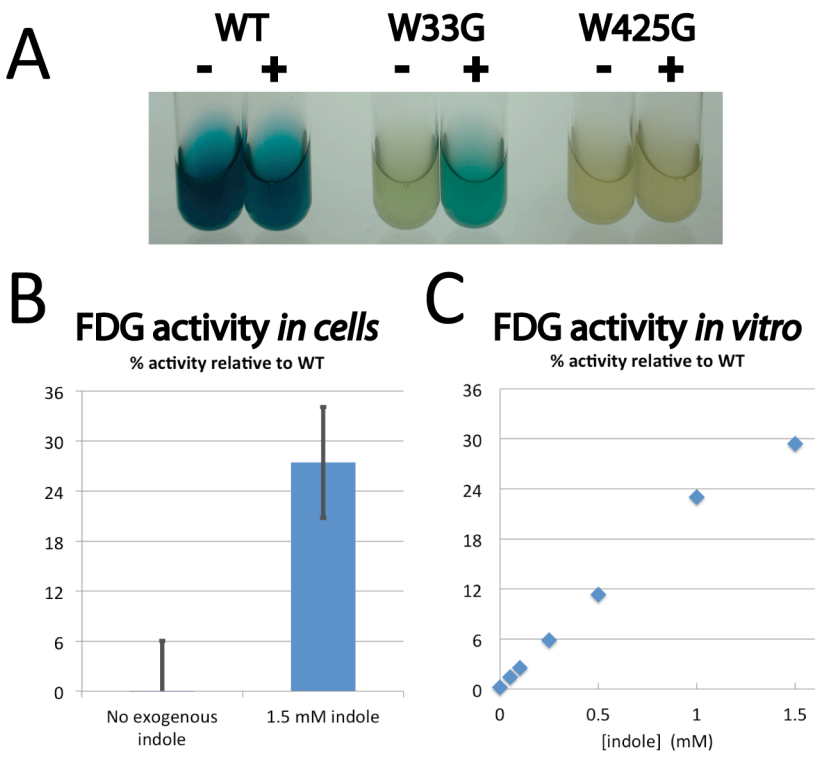
A
WT
- +
W33G
- +
W425G
- +
B
FDG activity in cells
% activity relative to WT
36
30
24
18
12
6
0
No exogenous indole
1.5 mM indole
C
FDG activity in vitro
% activity relative to WT
36
30
24
18
12
6
0
0
0.5
1
1.5
[indole] (mM)

Figure 4

Supplementary Figure Captions

Figure S1: Indole-dependence of activity *in vitro* using three β-gly constructs: wild type, W33G, and W425G. Wild type activity does not depend on indole, W33G activity is indole dependent, and W425G is inactive even upon addition of indole. Product formation was measured spectrophotometrically using 4 mM 2-nitrophenyl-β-D-galactopyranoside as substrate.

Figure S2: The simple allosteric kinetic scheme leading to the rate equation used in **Figure 2c**.

Figure S3: A difference electron density (Fo-Fc) map contoured at 3σ from holo β-gly W33G chain B before inserting the indole (the equivalent map for chain A is shown in **Figure 3c**). Electron density is consistent with indole and was modeled accordingly (gray) at a location overlapping Trp33 in the wild type enzyme (pale orange). The orientation of the indole cannot be unambiguously determined at this resolution.

Figure S4: Fractionation scheme used to distinguish between extracellular and intracellular activity. The absence of β-glycosidase activity in growth media after cells have been removed, even upon supplementing with indole, indicates that rescue occurs inside cells. While the rapid transport of indole across the *E. coli* cell membrane has been previously demonstrated [38], the *lack* of β-gly W33G activity in the absence of indole-supplemented media was unexpected. Indole is a critical signaling molecule in *E. coli* and a number of other bacterial species [39], yet under these

experimental conditions we do not observe β-gly W33G activation by either endogenous indole or any other structurally similar metabolites. While our inability to detect low levels of β-gly W33G activation by natural metabolites may be limited by the sensitivity of this crude biomolecular sensor, this observation may alternatively be explained by careful indole regulation by *E. coli*. Indeed, in separate experiments we find that under conditions of stress, considerably higher background activity can be observed.

Supplementary Figures

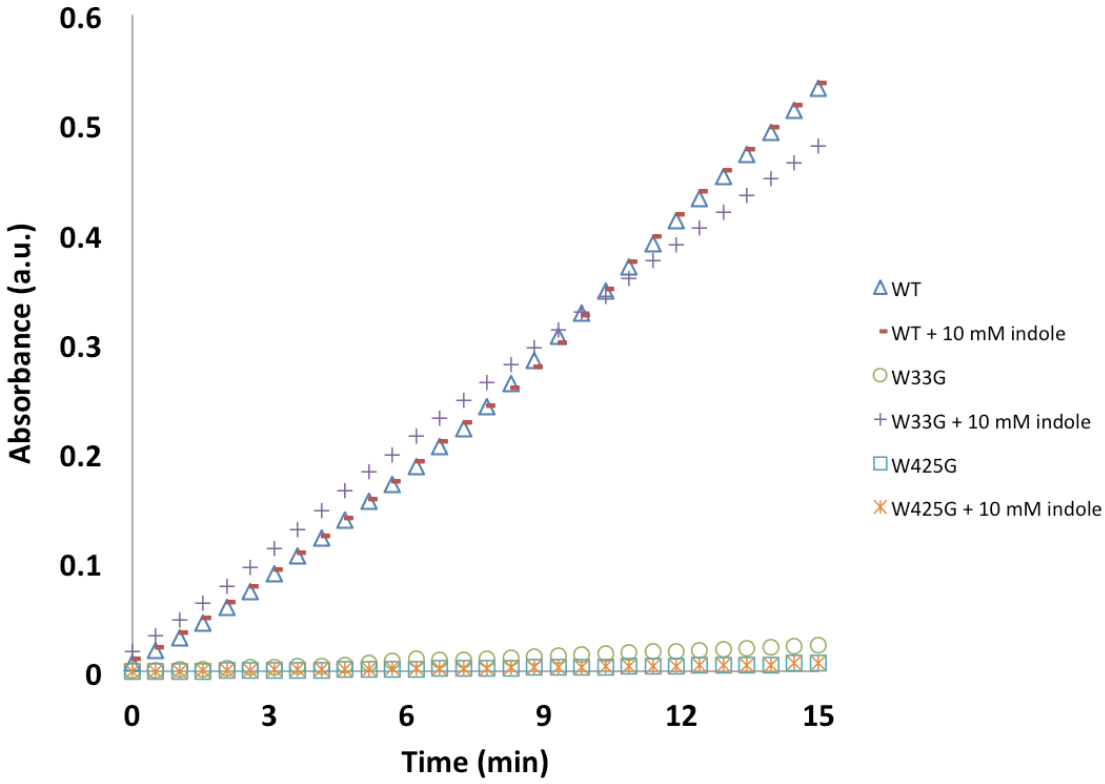

Figure S1

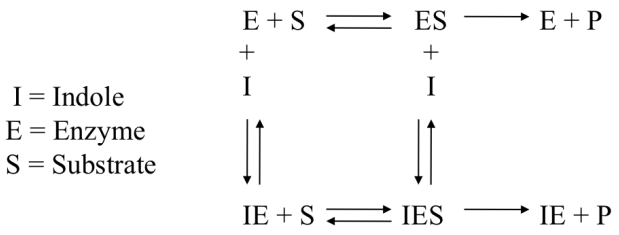

Figure S2

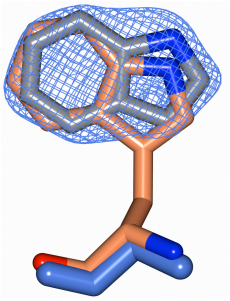

Figure S3

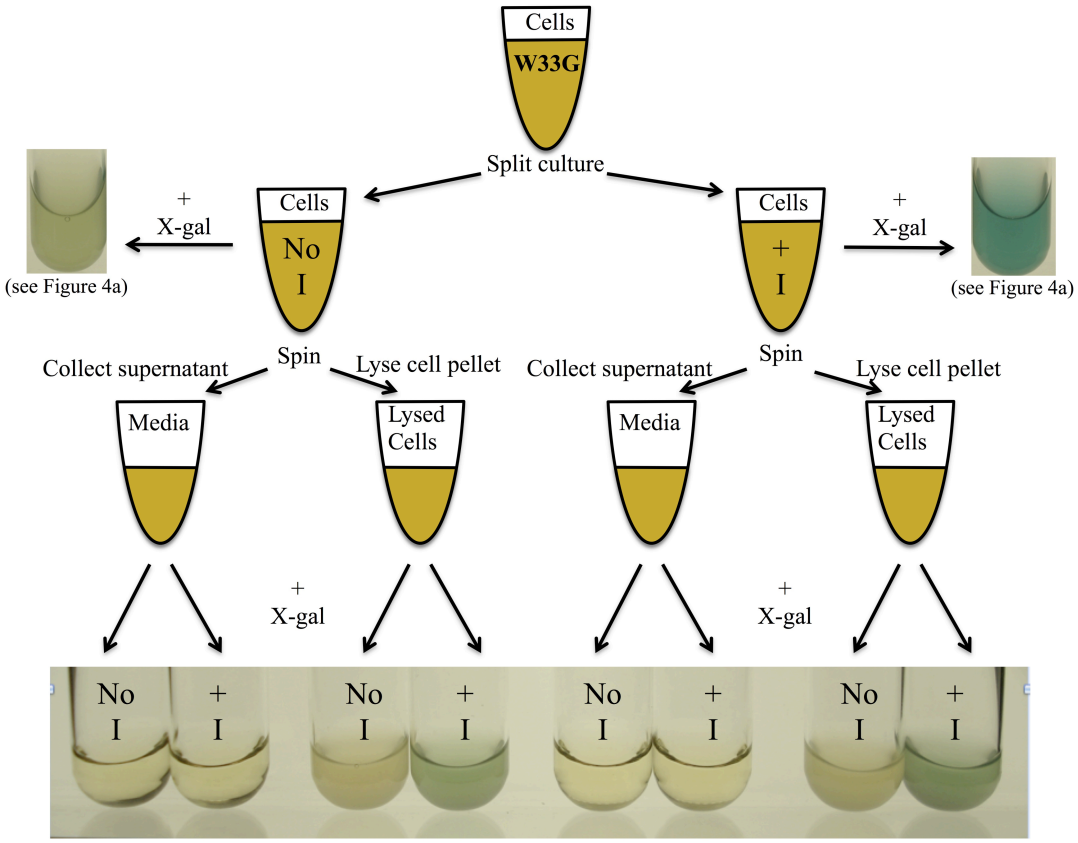
Cells
W33G
Split culture
+
X-gal
(see Figure 4a)
Cells
No
I
Cells
+
I
+
X-gal
(see Figure 4a)
Collect supernatant
Spin
Lyse cell pellet
Collect supernatant
Spin
Lyse cell pellet
Media
Lysed
Cells
Media
Lysed
Cells
+
X-gal
+
X-gal
No
I
+
I
No
I
+
I
No
I
+
I
No
I
+
I

Figure S4

Supplementary Table

Table S1: Crystallographic data for apo β-gly W33G refined to 1.70 Å resolution and for indole-bound β-gly W33G refined to 1.75 Å resolution.

	Apo W33G β-gly	Indole Bound W33G β-gly
Data Collection		
Unit-cell parameters (Å, °)	a = 167.70, c = 96.40	a = 167.54, c = 95.23
Space group	$P3_121$	$P3_121$
Resolution (Å)*	50.00-1.70 (1.79-1.70)	50.00-1.75 (1.84-1.75)
Wavelength (Å)	1.0000	1.0000
Temperature (K)	100	100
Observed reflections	1,150,966	1,029,278
Unique reflections	170,417	153,873
<I/σ(I)>*	12.6 (2.2)	14.4 (2.4)
Completeness (%)*	99.9 (100)	99.8 (100)
Multiplicity*	6.8 (7.1)	6.7 (6.9)
R_{merge} (%)*,#	8.8 (62.7)	7.9 (64.6)

R_{meas} *,^	9.5 (67.7)	8.6 (69.9)
R_{pim} *,^	3.7 (25.4)	3.3 (26.5)
Refinement		
Resolution (Å)	34.16-1.70	33.93-1.75
Reflections (working/test)	161,617 / 8,526	145,752 / 7,697
R_{factor} / R_{free} (%)&	16.8 / 18.6	17.4 / 19.1
No. of atoms (protein / Cl^- / MPD / Tris / indole / water)	7,965 / 6 / 48 / 16 / - / 719	7,981 / 6 / 16 / - / 18 / 615
Model Quality		
R.m.s deviations		
Bond lengths (Å)	0.010	0.010
Bond angles (°)	1.224	1.211
Average *B* factor (Å^2)		
All Atoms	21.6	25.0
Protein	20.5	24.2
Indole	-	34.4

Cl^-	26.4	31.7
MPD	34.7	40.1
Tris	30.0	-
Water	32.6	34.4
Coordinate error based on Maximum Likelihood (Å)	0.24	0.33
Ramachandran Plot		
Favored (%)	98.0	98.3
Allowed (%)	1.7	1.5

*Values in parenthesis are for the highest resolution shell.

$R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, 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.

& $R_{factor} = \Sigma_{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \Sigma_{hkl} |F_{obs}(hkl)|$; R_{free} 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} , [16].

R_{pim} = precision-indicating (multiplicity-weighted) R_{merge} , [40,41].