Published in final edited form as:
J Am Chem Soc. 2013 December 18; 135(50): 18840–18849. doi:10.1021/ja407644b.

The designability of protein switches by chemical rescue of structure: mechanisms of inactivation and reactivation

Yan Xia[1], **Nina DiPrimio**[2], **Theodore R. Keppel**[3], **Binh Vo**[1], **Keith Fraser**[4], **Kevin P. Battaile**[5], **Chet Egan**[1], **Christopher Bystroff**[4], **Scott Lovell**[6], **David D. Weis**[3], **J. Christopher Anderson**[2], and **John Karanicolas**[1,7,*]

[1]Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045

[2]Department of Bioengineering, California Institute for Quantitative Biological Research, University of California, Berkeley, CA 94720, USA

[3]Department of Chemistry, University of Kansas, Lawrence, KS 66045

[4]Department of Biology, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180

[5]IMCA-CAT, Hauptman-Woodward Medical Research Institute, 9700 South Cass Avenue, Building 435A, Argonne, Illinois 60439

[6]Protein Structure Laboratory, University of Kansas, Lawrence, KS 66045

[7]Center for Bioinformatics, University of Kansas, Lawrence, KS 66045

Abstract

The ability to selectively activate function of particular proteins via pharmacological agents is a longstanding goal in chemical biology. Recently, we reported an approach for designing a *de novo* allosteric effector site directly into the catalytic domain of an enzyme. This approach is distinct from traditional chemical rescue of enzymes in that it relies on disruption and restoration of structure, rather than active site chemistry, as a means to achieve modulate function. However, rationally identifying analogous *de novo* binding sites in other enzymes represents a key challenge for extending this approach to introduce allosteric control into other enzymes. Here we show that mutation sites leading to protein inactivation via tryptophan-to-glycine substitution and allowing (partial) reactivation by the subsequent addition of indole are remarkably frequent. Through a suite of methods including a cell-based reporter assay, computational structure prediction and energetic analysis, fluorescence studies, enzymology, pulse proteolysis, x-ray crystallography and hydrogen-deuterium mass spectrometry we find that these switchable proteins are most commonly

*To whom correspondence should be addressed. johnk@ku.edu, 785-864-8298.

Associated Content

Supporting Information

A complete description of experimental methods and procedures. Table S1 containing a summary of homodimeric *E. coli* proteins included in the reporter gene assay. Table S2 containing data and computational analysis generated from the cI reporter assay. Table S3 containing crystallographic data for +5 GFP W57G and W57A. Figure S1 showing dose-response of yeaZ W123G in the cI assay. Figure S2 showing the distribution of energies observed in simulations of each protein in the cI reporter assay. Figure S3 showing fluoresence intensities for +5 GFP W57A. Figure S4 showing excitation and emission spectra for +5 GFP W57A. Figure S5 showing dose-response for indole rescue of +5 GFP W57G fluoresence intensity. Figure S6 showing a difference electron density maps from +5 GFP W57G and W57A. Figure S7 showing a comparison of the crystal structures of +5 GFP wild-type, W57G, and W57A. Figure S8 showing the location of a water molecule occupying the cavity created in +5 GFP by the W57A mutation. Figure S9 showing the uncropped gel used in the pulse proteolysis experiment and a PMSF control gel. Figure S10 showing peptide coverage of β-gluc in the hydrogen/deuterium (H/D) exchange experiments. Figure S11 showing representative deuterium uptake curves. Figure S12 showing changes in deuterium uptake for individual peptides. Figure S13 showing a comparison of deuterium uptake in the rescued β-gluc relative to the wild-type enzyme. This material is available free of charge via the Internet at http://pubs.acs.org.

modulated *indirectly*, through control of protein stability. Addition of indole in these cases rescues activity not by reverting a discrete conformational change, as we had observed in the sole previously reported example, but rather rescues activity by restoring protein stability. This important finding will dramatically impact the design of future switches and sensors built by this approach, since evaluating stability differences associated with cavity-forming mutations is a far more tractable task than predicting allosteric conformational changes. By analogy to natural signaling systems, the insights from this study further raise the exciting prospect of modulating stability to design optimal recognition properties into future *de novo* switches and sensors built through chemical rescue of structure.

Introduction

Important advances in cell biology have been enabled through the ability to selectively activate proteins involved in key processes[1–7]. We recently described an approach for introducing allosteric control into enzymes via a strategy termed “chemical rescue of structure”[8]. This strategy entails introducing one or more cavity-forming mutations into a protein core at “buttressing” locations, i.e. where specific sidechains are critical for maintaining the structural integrity of the active site. Deletion of these “buttressing” residues leads to distortion of the active site geometry, and accordingly loss of enzyme activity. The subsequent addition of an exogenous compound that matches the deleted moiety is then expected to restore the “buttress” by binding in the cavity, and thus restore protein structure and activity.

Our previous studies[8] focused on β-glycosidase from *S. solfataricus* as a model enzyme. We introduced a tryptophan-to-glycine mutation (W33G) at a site close to (but distinct from) the active site, and found the ratio k_{cat}/K_m for this mutant to be about 730-fold worse than that of the wild-type enzyme. Upon solving the crystal structure of this mutant, we found that a very local conformational change distinguished it from the wild-type structure: a single nearby active site residue had shifted away from the active site to fill the cavity produced by the mutation. The change in position of this active site residue led to a loss of contact with the substrate, explaining the loss of function. We then found that exogenous indole could be used to completely restore activity to the mutant, with both k_{cat} and K_m reaching the corresponding values of the wild-type enzyme. The crystal structure of the mutant enzyme in complex with indole revealed that indole occupied exactly the cavity created by the mutation. This in turn perfectly restored the active site geometry, explaining the complete rescue of enzyme activity.

In contrast to chemical rescue of structure, most approaches for building ligand-dependent activity into enzymes have involved fusing a gene encoding some naturally-occurring allosteric “binding domain” (for the desired ligand) into a gene encoding some naturally-occurring “output domain” (for the desired activity)[9]. By using screens or selections to sift through the large number of potential insertion points and linkers, these fusions of existing protein domains have led to a variety of synthetic “switchable” proteins that are activated through allostery by the binding of an effector ligand[6,10–14]. The chemical rescue of structure approach is unique in that it introduces a ligand-binding site *directly* into the “output domain,” rather than rely on allosteric coupling to a separate “binding domain.” This alleviates the need for a naturally-occurring allosteric binding domain as a starting point, but instead requires that ligand binding alters *intradomain* function.

In the β-glycosidase example described above, the structural consequences of the cavity-forming mutation were indeed transduced to the active site, leading to loss and subsequent rescue of function. However, identifying cavity-forming mutations that induce analogous

conformational changes in other proteins represents a key challenge in building further *de novo* switches and sensors by chemical rescue of structure. Here, we seek to explore the general considerations that make this approach possible. In particular, we aim to address the following questions: How frequently does a single W→G cavity-forming mutation induce loss of function? How might one select sites that will lead to protein inactivation and reactivation by indole? And most importantly, must we explicitly tackle the challenge of modeling conformational changes resulting from cavity-forming mutations in order to predict sites at which chemical rescue of structure may be applied?

Results

Reporter gene assay for loss of function and indole rescue

To explore the frequency at which a W→G mutation leads to loss of function, we developed a reporter gene assay to monitor the loss and rescue of protein homodimerization *in vivo*. As a starting point we used the cI repressor from λ phage, which is comprised of an N-terminal DNA-binding domain and a C-terminal dimerization domain[15]. Upon homodimerization induced by the C-terminal domain, the N-terminal domain recognizes a pR promoter to repress downstream gene transcription[16]. To explore homodimerization in several different proteins, we created chimeras by separately fusing each protein to the N-terminal domain of cI. We expressed each chimera in *E. coli* that harbor a GFP gene under control of a pR promoter, and monitored GFP fluorescence in these cell cultures. By coupling the target construct dimerization (“function”) to transcriptional repression, this assay provides a straightforward readout of the protein’s oligomeric state (Figure 1A).

We applied this assay to test a total of 14 W→G mutations in three separate functionally unrelated *E.coli* genes that encode homodimeric proteins with available crystal structures: yeaZ, orn, and tadA (Table S1). As controls we used the reporter gene plasmid without cI repressor to monitor GFP fluorescence in the absence of repression (high RFUs), and we used the wild-type cI repressor to estimate the expected maximal repression (low RFUs); neither is strongly indole dependent. Chimeras produced by replacing the C-terminal domain of cI with any of the three wild-type homodimeric proteins led to repression comparable to that of the intact full-length wild-type cI repressor (Figure 1B, Table S2).

Upon introducing W→G mutations into these genes, we found that at least half disrupted repression of the GFP gene (Figure 1B, Table S2). The extent of repression from these mutants varied broadly: for example, yeaZ W134G and orn W9G had fluorescence intensities 96-fold and 59-fold higher than their wild-type counterparts (Table S2). In contrast, other mutants, such as yeaZ W169G and tadA W34G, maintained repressor activity nearly equivalent to that of their wild-type counterparts.

Subsequent addition of 1 mM indole to the cell cultures appeared to rescue repression in a number of cases: for example, yeaZ W123G, yeaZ W134G, yeaZ W159G, and orn W9G (Figure 1B, Table S2). For the mutant showing the greatest indole-induced relative difference in repression, yeaZ W123G, we further found that this enhanced repression responded smoothly to the concentration of indole (Figure S1). Though these results suggest that indole may restore dimerization in these mutants, the addition of indole did not result in complete repression of fluorescence back to the wild-type levels, most likely because higher concentrations of indole may be required for complete rescue[8]. Furthermore, despite the unchanged the fluorescence levels of reporter plasmid alone and wild-type protein chimeras upon addition of indole (Figure 1B, Table S2), we also cannot fully rule out the possibility that indole may cause the observed decrease in fluorescence through some other unrelated mechanism, such as unanticipated alterations in *E. coli* metabolism.

While this experiment does not explicitly normalize for possible changes in expression levels of our chimeric repressors, the observed differences in the behavior of W→G single-point mutants within the same construct are unlikely to be attributable to altered expression levels. To further investigate how structural changes upon incorporation of a W→G mutation may lead to inactivation in this experiment, we turned to simulation studies of these protein variants.

Structural analysis of mutations affecting dimerization

In order to develop a structure-based approach that would allow us to identify which tryptophan sidechains would lead to loss of function when mutated to glycine, we first labeled each tryptophan sidechain as "buttressing" (with respect to the dimer interface) or "not buttressing." Sites were labeled as "buttressing" if mutation to glycine led to at least 6-fold loss of repression in the cI reporter assay; 7 of the 14 mutation sites met this criterion (Table S2). We note that each of the proteins included in the cI assay has a different fold, and that the mutation sites are dispersed across each protein (Figure 2A).

On the basis of our studies of indole rescue in β-glycosidase[8], we expected that protein inactivation would again result from an allosteric conformational change that coupled disruption at the mutation site to distortion at the functional site (in this case, the dimer interface). We further reasoned that such allosteric conformational changes—if not explicitly evolved or designed—would be more likely to occur locally than over long distances through the protein. As a first indirect test of this hypothesis, we therefore computed the distance of each mutation site to the dimer interface (see Supporting Information), with the expectation that the mutations closest to the dimer interface would most frequently be those producing loss of repression in the cI assay.

To evaluate the accuracy of this approach for predicting the effect of these cavity-forming mutations, we turned to receiver operating characteristic (ROC) analysis. Using the distance to the interface as our predictor, we plotted the fraction of true positives identified in our set (sites that are "buttressing" and are correctly classified as such) versus the fraction of false positives (non-buttressing sites that are incorrectly classified as "buttressing"), for increasing values of the discrimination distance threshold. Using this analysis, the curve for a perfect predictor will rise vertically to the upper left corner of the plot; in contrast, a method that makes predictions at random have a curve that approximately follows the diagonal (*red dashed line*). While mutations to either of the two tryptophan sites closest to the dimer interface indeed led to loss of repression (orn W9G and orn W143G), this approach failed to readily identify the other five buttressing sites (Figure 2B); overall, this predictor performed essentially as a random predictor.

To further explore the hypothesis that disruption at the mutation site could be coupled to distortion of the dimer interface through some distinct conformational change, we used structure prediction tools in the Rosetta macromolecular modeling suite[17–19] to probe the structural consequences of each mutation. We treated prediction of each mutant structure as a comparative modeling task, using the crystal structure of the wild-type dimer as a template for refinement (see *Methods*). For each of the resulting output structures, we evaluated the interaction energy between the two subunits and compared it to the corresponding energy in the wild-type structure: our hypothesis was that specific structural changes resulting from mutations at buttressing residues might lead to disruption of interactions in the protein-protein interface. However, this approach also performed essentially as a random predictor (Figure 2C), suggesting that direct consideration of interface energetics—predicated on building structural models from the wild-type template—was incapable of explaining why certain W→G mutations led to loss of repression while others did not.

We next surmised that perhaps these drastic cavity-forming mutations had destabilized the protein to the point of inducing local or global unfolding[20–23], in which case the crystal structure of the wild-type dimer may not prove to be suitable template structure prediction. Starting from the premise that the likelihood of a long-range allosteric conformational change in response to an arbitrary mutation is rare, we postulated that a protein could respond in three other ways to a W→G mutation: absorb the energetic cost of maintaining a cavity in the hydrophobic core of the protein, undergo local collapse of nearby structure to minimize the occupied volume in the core, or unfold. Given that the structural response to mutations that decrease sidechain volume can vary substantially depending on context[20], we returned to the comparative models we had previously built. Using these models, in which local reorganization may have been captured by our refinement protocol, we used Rosetta to estimate the stability difference of each mutant (dimeric) protein relative to the corresponding wild-type dimer (see *Methods*).

In stark contrast to the previous approaches, the estimated stability difference proved to be an outstanding predictor of which W→G mutations would lead to loss of repression in the cI assay (Figure 2D). We further note that the difference in average energy associated with each mutant came not from a small number of outlying conformations, but rather from a systematic shift in energy over the entire ensemble (Figure S2); while there is variation from averaging over the set of conformations, the nature of these differences thus highlights the robustness of this method for estimating stability differences.

In addition being a powerful binary classifier, the estimated stability difference also gave *quantitative* correlation with the relative fluorescence measured in the cI reporter assay, with Spearman rank correlation coefficient $\rho=0.69$, a statistically significant non-zero value ($p < 0.008$). The excellent predictive power of this approach supports the hypothesis that the loss of dimerization in the cI repressor assay was caused by loss of protein stability rather than a discrete conformational change. To test this novel mechanism for inactivation and rescue, we next turned to direct biochemical characterization.

Mechanism of inactivation and rescue in +5 GFP

Due to the inherent challenges associated with the biochemical and structural characterization of homodimers, we elected to explore whether the same stability-mediated mechanism of inactivation and rescue occurred in a model system more naturally amenable to these *in vitro* techniques. We selected +5 GFP for these studies, a variant of “superfolder” GFP[24]. Like most GFP constructs, +5 GFP folds into a β-barrel harboring a single tryptophan residue (Trp57) on the central helix, 10 Å from the chromophore[24]. Simulations analogous to those described above gave an estimated stability difference of 4.5 Rosetta energy units associated with this W57G mutation; this value nearly, but not quite, reaches the threshold of 5.0 over which we regularly observed loss of function in the cI reporter assay (Table S2).

We measured the fluorescence intensity of wild-type +5 GFP and its W57G mutant, and found that deletion of this tryptophan sidechain reduced the fluorescence intensity by 50% (Figure 3A). While addition of 1 mM indole led to a slight decrease in fluorescence intensity for the wild-type protein, indole instead *rescued* fluorescence in the W57G mutant, back to 63% of the wild-type value (the difference in fluorescence intensity upon addition of indole to +5 GFP W57G is statistically significant, $p < 0.001$ using Welch’s t-test). Rescue of W57G fluorescence by indole increases in a dose-dependent manner (Figure S5).

It is well established that slight structural rearrangements close to the GFP chromophore can lead to dramatic spectral differences[25,26]; the fluorescence properties can thus serve as a sensitive readout of the local environment surrounding the chromophore. We therefore

carried out excitation and emission wavelength scans for both GFP constructs (Figure 3B). The shapes of the wild-type and W57G spectra are identical, notwithstanding a 46% decrease in intensity upon mutation (consistent with Figure 3A). The addition of 1 mM indole did not change either curve shape, save the same intensity differences observed previously (Figure 3A). Collectively, the lack of peak shifts or additional peaks in these spectra suggests that the partial inactivation and rescue we observed was not coupled to reorganization of the packing around the chromophore.

Based on the unchanged excitation and emission maxima, we formulated the hypothesis that in the absence of indole, +5 GFP W57G populates two states. The first, comprised of 46–50% of the population, is characterized by a conformation very similar to that of wild-type +5 GFP and accounts for the native-like excitation and emission spectra. The second state, accounting for the remaining 50–54% of the population, may be partly unfolded or have changes in conformational dynamics that disrupt the chromophore and result in loss of fluorescence.

To test this hypothesis, we solved the crystal structure of +5 GFP W57G to 1.6 Å resolution (Table S3). While it was somewhat surprising to obtain crystals from the heterogeneous population we anticipated, we postulate that the (non-equilibrium) process of crystallization allowed us to capture the native-like (fluorescent) state. Accordingly, our solved structure of +5 GFP W57G closely resembles the structure of wild-type +5 GFP previously determined[24], with overall Cα RMSD of 0.84 Å (229 residues), Cα RMSD for residues within 4 Å of the chromophore of 0.25 Å (21 residues), and no structural differences evident in response to the mutation (Figure 3C). We also found that +5 GFP W57A exhibited similar fluorescence properties as +5 GFP W57G including rescue by indole (Figures S3, S4), and yielded crystals that diffracted to 1.1 Å resolution (Table S3). Like +5 GFP W57G, the crystal structure of +5 GFP W57A showed no structural differences relative to the wild-type structure, including the backbone at the site of the mutation (Figures S6, S7). Interestingly, the +5 GFP W57A structure revealed a water molecule located exactly at the position previously occupied by the indole nitrogen of Trp57, recapitulating the hydrogen bond to a nearby aspartate observed in the wild-type structure (Figure S8). While both the W57G and the W57A structures contain a large cavity previously filled by the tryptophan sidechain, this cavity is neither completely occluded from solvent nor completely hydrophobic; this makes it unsurprising that water occupies the space vacated by either mutation[27].

With this evidence that fluorescence in +5 GFP W57G derives from a species having essentially the wild-type structure, we next sought evidence for an alternate state comprising the remainder of the population. To probe for such a state we carried out a pulse proteolysis experiment, incubating either of wild-type +5 GFP or its W57G mutant with subtilisin. We found that while the folded native structure of wild-type +5 GFP renders it largely protected from proteolysis, the W57G mutant is extensively digested almost immediately (Figure 3D). We further found that inclusion of a protease inhibitor (PMSF) in the reaction prevents loss of +5 GFP W57G, while DMSO (used as a vehicle for PMSF) does not (Figure S9). The fact that PMSF prevents the disappearance of W57G +5 GFP serves to confirm that indeed proteolysis is responsible, and not some other process such as aggregation. The susceptibility of +5 GFP W57G to proteolysis supports the hypothesis that in addition to a state that strongly resembles wild-type +5 GFP, this mutant also populates a state in which subtilisin cleavage sites are more exposed than in its native-like (fluorescent) conformation. While we speculate that addition of indole would confer enhanced subtilisin resistance to W57G +5 GFP, we found through separate control experiments (not shown) that indole itself inhibits this protease directly; this made it impossible to test for indole rescue of W57G +5 GFP subtilisin resistance.

Collectively, these observations point to a model in which incorporation of the W57G mutation into +5 GFP induces unfolding or enhanced fluctuations in a subset of the population (loss of fluorescence intensity), followed by a shift in this population back to the native-like state upon addition of indole (rescue of fluorescence intensity). This model is qualitatively distinct from the mechanism of inactivation and rescue we observed in our characterization of β-gly W33G[8].

Mechanism of inactivation and rescue in β-glucuronidase

Motivated by this stability-mediated model for inactivation and rescue of +5 GFP W57G, we returned to the *E. coli* β-glucuronidase (β-gluc) W492G mutant described previously[8]. We had characterized this enzyme only in passing as part of our initial studies of indole rescue, showing that indole could be used to partially restore activity to this mutant in a dose-dependent manner. Though the structure of the wild-type enzyme has been solved via X-ray crystallography[28], we found that the W492G mutant was not amenable to crystallization. We further applied the Rosetta refinement tools[18,19] to build comparative models of the W492G mutant, with the structure of the wild-type enzyme as a template; none of these models, however, included a conformational change linking the mutation site to the active site. In the absence of any structural insights we were, at the time, unable to explain the basis for the loss of enzyme activity due to this mutation[8], particularly given that the mutation site lies 13 Å from the enzyme active site.

In light of the studies we reported above, we formulated the hypothesis that the indole-dependent activity of β-gluc W492G may also be modulated by enhanced fluctuations or local/global unfolding, which are then reverted upon addition of indole. This hypothesis could explain our inability to form crystals of the W492G mutant, and also our inability to build a compelling model of the structure of this protein. This hypothesis was further supported by the stability difference of 6.5 Rosetta energy units estimated for this W57G mutation, above the threshold of 5.0 that proved predictive in the cI reporter assay (Table S2).

To directly test this hypothesis, we used hydrogen/deuterium (H/D) exchange experiments to probe local fluctuation events in the protein. Upon incubation with deuterium-containing solvent, amides that are not strongly hydrogen bonded are more rapidly isotopically labeled than amides involved in intramolecular hydrogen bonds[29–32]. Consequently, hydrogen–deuterium exchange allows us to localize conformational differences between β-gluc variants or upon addition of indole. The large size of this enzyme precluded straightforward residue-level localization of deuterium exchange information via NMR. For this reason, we instead quenched the exchange reaction, used pepsin to digest the protein, and then quantified the extent of exchange for each peptide fragment via mass spectrometry[33] (see *Methods*). A total of 147 peptides, of average length 13 residues, collectively covered 82% of the whole protein sequence excluding proline residues (Figure S10); this included good coverage near the active site and the mutation site (Figure 4A), and extensive overlap in many regions. We separately incubated wild-type β-gluc and the W492G variant in deuterium-containing buffer, both in the absence of indole and in the presence of 5 mM indole. Aliquots at multiple time points were digested and analyzed by mass spectrometry to determine the degree to which the protein environment conferred protection from exchange at specific regions of the protein (see Supporting Information, Figure S11).

Relative to the wild-type enzyme, a number of segments from the W492G variant showed enhanced deuterium uptake, corresponding to less protection by the protein environment. Upon addition of indole, many of the *same* peptides exhibited decreased deuterium uptake, suggesting that indole reverted the effect of this mutation (Figure S12). To allow direct comparison between peptides of different sizes, we calculated for each peptide the

"normalized deuterium difference", *NDD*, defined as the difference in peptide mass increase per exchangeable amide hydrogen, averaged over all time points (see Supporting Information). To further localize the effect of mutation and indole rescue, we then returned to the mapping of each peptide to the protein sequence. At every position in the protein sequence, we assigned the normalized deuterium difference for the residue as the average *NDD* value for all peptides covering its position in the sequence. While this does allow calculation of an *NDD* value for all residues covered by at least one peptide, we note that these *NDD* values are not truly residue-resolved, since each peptide represents information integrated from adjacent residues as well as the residue of interest.

Relative to the wild-type enzyme, we observe enhanced deuterium uptake in the W492G mutant that is localized to specific regions of the protein sequence (Figure 4B). Addition of indole to the wild-type enzyme does not result in appreciable changes in deuterium uptake (Figure 4C); in contrast, addition of indole to β-gluc W492G leads to protection against deuterium uptake (Figure 4D). We further note that most of the regions in this mutant that exhibit increased protection upon addition of indole are the *same* as those that showed enhanced deuterium uptake as a result of the mutation. Upon comparing deuterium uptake between wild-type β-gluc and W492G with 5 mM indole present in both, we find that indole does not change the pattern but slightly reduces the magnitude of the differences (Figure S13). Our observation that the enhanced deuterium uptake of the mutant is not fully abrogated by addition of 5 mM indole is not surprising, given that we previously observed only partial rescue of enzyme activity at this indole concentration[8].

Mapping *NDD* values to the structure of the wild-type enzyme reveals a cohesive picture of inactivation and rescue. First, introduction of the W492G mutation leads to less protection from deuterium uptake in a nearby region that includes two helices and several intervening loops, indicating that loss of function in response to this cavity-forming mutation occurs due to enhanced fluctuations or local unfolding (Figure 4E). Addition of indole then restores protection from deuterium uptake at the same regions (Figure 4F), suggesting rigidification or refolding of these regions around the indole. These changes induced by addition of indole (partially) shift the conformational ensemble back towards that of the wild-type enzyme, thus providing a structural explanation for the previously-unexplained (partial) rescue of enzyme activity[8].

Discussion

In our earlier work[8] we identified two examples of residues required for buttressing the nearby active site: removal (via cavity-forming mutation) of a sidechain playing this key role in maintaining the protein architecture results in collapse of the active site geometry, and thus loss of function. Our structural studies of β-gly W33G revealed a distinct conformational change induced by the cavity-forming mutation, which fortuitously transduced this disruption to the active site. Predicting the long-range effects of structural variations in general represents a very challenging problem[34–40], making it exceedingly unlikely that such predictions can be routinely used to introduce analogous mutations for building allosteric control into other proteins.

The systematic evaluation of a larger test set in our cI repressor assay (Figure 1) and the subsequent computational analysis (Figure 2), implied that protein structure and function could instead be modulated *indirectly*, through control of protein stability. In both examples for which we subsequently carried out detailed biochemical studies (Figure 3, 4), we found strong evidence pointing to enhanced fluctuations or unfolding resulting from destabilization as the mechanism underlying loss of function upon mutation. Accordingly, reactivation by

indole may occur not only by reversion of a discrete conformational change (as in β-gly W33G), but alternatively by rigidifying or refolding the protein to its active state.

It is also noteworthy that all of the proteins characterized here derive from mesophilic organisms, whereas the β-glycosidase we studied previously derives from a hyperthermophilic organism (*Sulfolobus solfataricus*). The extreme stability of β-gly may have rendered it resistant to unfolding, allowing it to instead respond to the cavity-forming mutation via the conformational change we described earlier[8]. In light of the results presented here, we expect that modulation of function via chemical rescue of structure will rely, for most proteins from mesophilic organisms, on stability-mediated mechanisms.

In order to build small-molecule dependence into a protein domain via chemical rescue of structure, the cavity-forming mutation must induce the protein to undergo a transition to some non-functional state (Figure 5); however, the precise details of this inactive state need not be explicitly designed. Attempting to rationally identify cavity-forming mutations to inactivate some protein of interest via a discrete conformational change would prove exceedingly challenging; on the other hand, evaluating the stability difference associated with cavity-forming mutations represents a far more tractable task. Accordingly, we expect that the insights offered here will immediately enable rational design of a variety of new protein switches that rely on activation by indole-induced protein stabilization, and will prove highly complementary to techniques that use inducible “degron” fusions to modulate degradation by the proteasome[3,41].

Natural systems make use of small-molecules to encode a broad range of signals, whose diversity is reflected in the wide variety of mechanisms that are used to transduce ligand binding into downstream activity[7]. These mechanisms range from discrete conformational changes[42,43], to population shifts[44–47], to induced folding[48–51]. Using chemical rescue of structure, we have already observed a similar range of mechanisms for recognition and activation in our set of designed protein switches.

The spectrum of available mechanisms for recognition enable natural systems to produce switches with vastly different properties. There are specific design advantages associated with using each distinct mechanism: these may include intrinsic differences in dynamic range[52], selectivity[53,54], kinetics[51,55], and the ability to modulate signals by altering cellular accumulation through resistance to proteolysis[50,56]. The precise functional requirements associated with responding to a particular stimulus can therefore be met, in part, by selecting a recognition mechanism that will confer the desired kinetics, sensitivity, and dynamic range[7]. The observation that our designed protein switches exhibit the same range of mechanisms for activation as do natural switches raises the exciting prospect that – by carefully selecting a protein host with appropriate structure and stability – we too can build the desired activation mechanism into our protein switches. This, in turn, may enable us to tune the specific properties of *de novo* switches built through chemical rescue of structure, and thus allow us to cater to the unique criteria presented in biological signaling and sensing applications.

Methods

cI repressor assay

We transformed the reporter vector into *E.coli* cells DIAL strain JI[57] to create the reporter strain, then transformed the expression plasmids for cI–target fusion proteins into the reporter strain. Cell cultures were split into equal aliquots, and indole was introduced into the media for a final concentration of 1 mM using DMSO as a vehicle. 1% DMSO alone was added to cell cultures not treated with indole. All experiments were carried out in

triplicate. Fluorescence intensities of GFP were detected spectrophotometrically. Further assay details are provided in Supporting Information.

Rosetta refinement protocol

We generated the starting structure of each W→G mutant by changing the amino acid identity at the mutation site and removing sidechain atoms using a text editor. We then performed 1000 independent simulations for each wild-type protein and W→G mutant in our set, using the "relax" protocol[17] in the Rosetta macromolecular modeling suite[18]. We used the average energy of the 100 best-scoring output structures (10% of those generated) to quantify the predicted interaction energy of the dimer interface or the relative stability of the protein. Differences in the interaction energy of the dimer interface or in the protein stability resulting from a mutation were estimated by subtracting the wild-type value from that of the mutant. The simulation protocol is further described in Supporting Information.

Fluorescence measurements of +5 GFP

Fluorescence intensity was measured on a Synergy2 BioTek plate reader with a 485/20 nm excitation filter and a 528/20 nm emission filter, using black 96-well plates. Ten replicates were assayed for each sample. Excitation/emission scans were carried out using a Cary Eclipse spectrophotometer (Varian) in a quartz cuvette.

All +5 GFP fluorescence studies (for wild-type, W57G, and W57A) were carried out using 9.5 μg/ml of protein with 5% DMSO and either no indole or 1 mM indole. Samples were incubated for 1 hour prior to reading. Buffer conditions are described in Supporting Information.

Crystal structures of +5 GFP mutants

Coordinates and structure factors for the crystal structures of +5 GFP W57G and W57A have been deposited with the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) with accession codes 4LQU and 4LQT, respectively. Crystallographic methods and data are presented in Supporting Information.

Pulse proteolysis of +5 GFP

Samples of +5 GFP (wild-type or W57G) at 0.9 mg/ml were incubated with increasing concentrations of subtilisin at 37°C. After 1 hour the proteolysis reaction was quenched by adding an equal volume of SDS-PAGE loading buffer and heating at 95°C for 15 min. Samples were analyzed by SDS-PAGE using Coomassie staining, and band intensities were quantified using ImageJ[58]. Further details are provided in Supporting Information.

Hydrogen–deuterium exchange for β-glucuronidase

To initiate deuterium labeling, 0.5 μg/μL of wild-type or W492G β-gluc was diluted 20-fold into D_2O buffer with and without 5 mM indole. The reaction mixture was incubated at 25°C for a time course ranging from 10 seconds to 12 hours, then was quenched using cold hydrochloric acid to pH 2.6. Quenched protein samples were immediately digested with pepsin. The peptides were separated by high-performance liquid chromatography and analyzed by time-of-flight mass spectrometry. A complete description of these methods is provided in Supporting Information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Andrea Bazzoli for critically reading the manuscript and offering helpful suggestions. We thank S. Jimmy Budiardjo and Katelyn Deckert for valuable discussions. We thank Susan Marqusee for suggesting the pulse proteolysis experiment and Ragul Gowthaman for assistance with statistical analysis. 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 National Center for Research Resources (5P20RR017708-09) (J.K.), the National Institute of General Medical Sciences (8P20GM103420-10) (J.K.), the National Science Foundation through TeraGrid allocation TG-MCB130049 (J.K.), the Human Frontier Science Program (J.K.), the University of Kansas General Research Fund allocation #2301390 (J.K.), and the Alfred P. Sloan Fellowship (J.K.).

References

1. Spencer DM, Wandless TJ, Schreiber SL, Crabtree GR. Science. 1993; 262:1019. [PubMed: 7694365]
2. Buskirk AR, Ong YC, Gartner ZJ, Liu DR. Proc Natl Acad Sci U S A. 2004; 101:10505. [PubMed: 15247421]
3. Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. Cell. 2006; 126:995. [PubMed: 16959577]
4. Pratt MR, Schwartz EC, Muir TW. Proc Natl Acad Sci U S A. 2007; 104:11209. [PubMed: 17563385]
5. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ, Ranganathan R. Science. 2008; 322:438. [PubMed: 18927392]
6. Karginov AV, Ding F, Kota P, Dokholyan NV, Hahn KM. Nat Biotechnol. 2010; 28:743. [PubMed: 20581846]
7. Ha JH, Loh SN. Chemistry. 2012; 18:7984. [PubMed: 22688954]
8. Deckert K, Budiardjo SJ, Brunner LC, Lovell S, Karanicolas J. J Am Chem Soc. 2012; 134:10055. [PubMed: 22655749]
9. Kanwar M, Wright RC, Date A, Tullman J, Ostermeier M. Methods Enzymol. 2013; 523:369. [PubMed: 23422439]
10. Guntas G, Mansell TJ, Kim JR, Ostermeier M. Proc Natl Acad Sci U S A. 2005; 102:11224. [PubMed: 16061816]
11. Stratton MM, Loh SN. Protein Sci. 2011; 20:19. [PubMed: 21064163]
12. Dueber JE, Yeh BJ, Chak K, Lim WA. Science. 2003; 301:1904. [PubMed: 14512628]
13. Meister GE, Joshi NS. Chembiochem. 2013; 14:1460. [PubMed: 23825049]
14. Dagliyan O, Shirvanyants D, Karginov AV, Ding F, Fee L, Chandrasekaran SN, Freisinger CM, Smolen GA, Huttenlocher A, Hahn KM, Dokholyan NV. Proc Natl Acad Sci U S A. 2013; 110:6800. [PubMed: 23569285]
15. Bell CE, Frescura P, Hochschild A, Lewis M. Cell. 2000; 101:801. [PubMed: 10892750]
16. Dodd IB, Shearwin KE, Egan JB. Curr Opin Genet Dev. 2005; 15:145. [PubMed: 15797197]
17. Qian B, Raman S, Das R, Bradley P, McCoy AJ, Read RJ, Baker D. Nature. 2007; 450:259. [PubMed: 17934447]
18. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. Methods Enzymol. 2011; 487:545. [PubMed: 21187238]
19. Raman S, Vernon R, Thompson J, Tyka M, Sadreyev R, Pei J, Kim D, Kellogg E, DiMaio F, Lange O, Kinch L, Sheffler W, Kim BH, Das R, Grishin NV, Baker D. Proteins. 2009; 77(Suppl 9):89. [PubMed: 19701941]
20. Eriksson AE, Baase WA, Wozniak JA, Matthews BW. Nature. 1992; 355:371. [PubMed: 1731252]

21. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW. Science. 1992; 255:178. [PubMed: 1553543]
22. Xu B, Hua QX, Nakagawa SH, Jia W, Chu YC, Katsoyannis PG, Weiss MA. Protein Sci. 2002; 11:104. [PubMed: 11742127]
23. Xu J, Baase WA, Baldwin E, Matthews BW. Protein Sci. 1998; 7:158. [PubMed: 9514271]
24. Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Nat Biotechnol. 2006; 24:79. [PubMed: 16369541]
25. Ehrig T, O'Kane DJ, Prendergast FG. FEBS Lett. 1995; 367:163. [PubMed: 7796912]
26. Palm GJ, Zdanov A, Gaitanaris GA, Stauber R, Pavlakis GN, Wlodawer A. Nat Struct Biol. 1997; 4:361. [PubMed: 9145105]
27. Matthews BW, Liu L. Protein Sci. 2009; 18:494. [PubMed: 19241368]
28. Wallace BD, Wang H, Lane KT, Scott JE, Orans J, Koo JS, Venkatesh M, Jobin C, Yeh LA, Mani S, Redinbo MR. Science. 2010; 330:831. [PubMed: 21051639]
29. Hvidt A, Linderstrom-Lang K. Biochim Biophys Acta. 1954; 14:574. [PubMed: 13198919]
30. Englander SW, Kallenbach NR. Q Rev Biophys. 1983; 16:521. [PubMed: 6204354]
31. Skinner JJ, Lim WK, Bedard S, Black BE, Englander SW. Protein Sci. 2012; 21:996. [PubMed: 22544544]
32. Skinner JJ, Lim WK, Bedard S, Black BE, Englander SW. Protein Sci. 2012; 21:987. [PubMed: 22544567]
33. Zhang Z, Smith DL. Protein Sci. 1993; 2:522. [PubMed: 8390883]
34. Machicado C, Bueno M, Sancho J. Protein Eng. 2002; 15:669. [PubMed: 12364581]
35. Demerdash ON, Daily MD, Mitchell JC. PLoS Comput Biol. 2009; 5:e1000531. [PubMed: 19816556]
36. Dixit A, Verkhivker GM. PLoS Comput Biol. 2011; 7:e1002179. [PubMed: 21998569]
37. Kidd BA, Baker D, Thomas WE. PLoS Comput Biol. 2009; 5:e1000484. [PubMed: 19714199]
38. Laine E, Goncalves C, Karst JC, Lesnard A, Rault S, Tang WJ, Malliavin TE, Ladant D, Blondel A. Proc Natl Acad Sci U S A. 2010; 107:11277. [PubMed: 20534570]
39. Huang W, Lu S, Huang Z, Liu X, Mou L, Luo Y, Zhao Y, Liu Y, Chen Z, Hou T, Zhang J. Bioinformatics. 2013; 29:2357. [PubMed: 23842804]
40. Goncearenco A, Mitternacht S, Yong T, Eisenhaber B, Eisenhaber F, Berezovsky IN. Nucleic Acids Res. 2013; 41:W266. [PubMed: 23737445]
41. Cho U, Zimmerman SM, Chen LC, Owen E, Kim JV, Kim SK, Wandless TJ. PLoS One. 2013; 8:e72393. [PubMed: 23991108]
42. Amemiya T, Koike R, Fuchigami S, Ikeguchi M, Kidera A. J Mol Biol. 2011; 408:568. [PubMed: 21376729]
43. Flores S, Echols N, Milburn D, Hespenheide B, Keating K, Lu J, Wells S, Yu EZ, Thorpe M, Gerstein M. Nucleic Acids Res. 2006; 34:D296. [PubMed: 16381870]
44. Pan H, Lee JC, Hilser VJ. Proc Natl Acad Sci U S A. 2000; 97:12020. [PubMed: 11035796]
45. Motlagh HN, Hilser VJ. Proc Natl Acad Sci U S A. 2012; 109:4134. [PubMed: 22388747]
46. Bruning JB, Parent AA, Gil G, Zhao M, Nowak J, Pace MC, Smith CL, Afonine PV, Adams PD, Katzenellenbogen JA, Nettles KW. Nat Chem Biol. 2010; 6:837. [PubMed: 20924370]
47. Wrabl JO, Gu J, Liu T, Schrank TP, Whitten ST, Hilser VJ. Biophys Chem. 2011; 159:129. [PubMed: 21684672]
48. Hilser VJ, Thompson EB. J Biol Chem. 2011; 286:39675. [PubMed: 21937423]
49. Motlagh HN, Li J, Thompson EB, Hilser VJ. Biochem Soc Trans. 2012; 40:975. [PubMed: 22988850]
50. Babu MM, van der Lee R, de Groot NS, Gsponer J. Curr Opin Struct Biol. 2011; 21:432. [PubMed: 21514144]
51. Wright PE, Dyson HJ. Curr Opin Struct Biol. 2009; 19:31. [PubMed: 19157855]
52. Vallee-Belisle A, Ricci F, Plaxco KW. J Am Chem Soc. 2012; 134:2876. [PubMed: 22239688]
53. Uversky VN. Biochim Biophys Acta. 2011; 1814:693. [PubMed: 21440685]

54. Hazy E, Tompa P. Chemphyschem. 2009; 10:1415. [PubMed: 19462392]
55. Huang Y, Liu Z. J Mol Biol. 2009; 393:1143. [PubMed: 19747922]
56. Choi JH, San A, Ostermeier M. Protein Sci. 2013; 22:475. [PubMed: 23400970]
57. Kittleson JT, Cheung S, Anderson JC. J Biol Eng. 2011; 5:10. [PubMed: 21787416]
58. Schneider CA, Rasband WS, Eliceiri KW. Nat Methods. 2012; 9:671. [PubMed: 22930834]

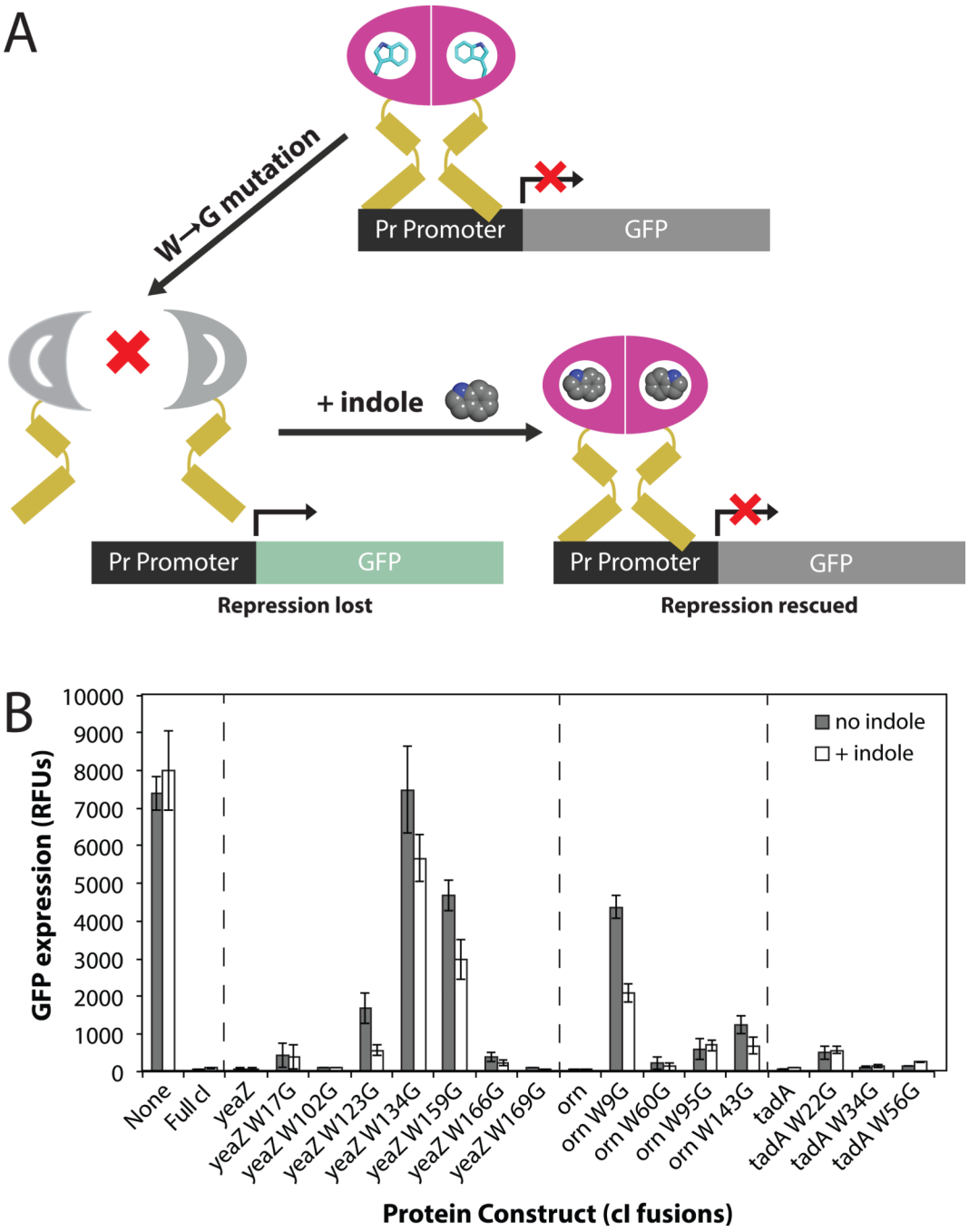

Figure 1. Loss of function and indole rescue in a cell-based assay

(A) Schematic of the cI repressor assay. Various homodimeric proteins (*pink*) are fused to the DNA-binding domain of cI repressor (*yellow*), enabling binding at the Pr promoter and repression of the GFP gene. A W→G mutation that disrupts dimerization will lead to loss of repression, and thus increased expression of GFP. If the subsequent addition of indole rescues dimerization, repression will be restored and GFP expression will decrease. ***(B)*** Effect of individual W→G mutations, and the subsequent addition of 1 mM indole, determined by GFP expression in the cI repressor assay (relative fluorescence units, *RFU*). More than half of the mutations disrupted repression of the GFP gene; repression could then

be partially rescued by addition of indole in a number of cases. Notable examples exhibiting loss of repression and subsequent rescue include yeaZ W123G, yeaZ W134G, yeaZ W159G, and orn W9G. Figure S1 shows the effect of indole concentration on rescue of yeaZ W123G. (Supporting Information).

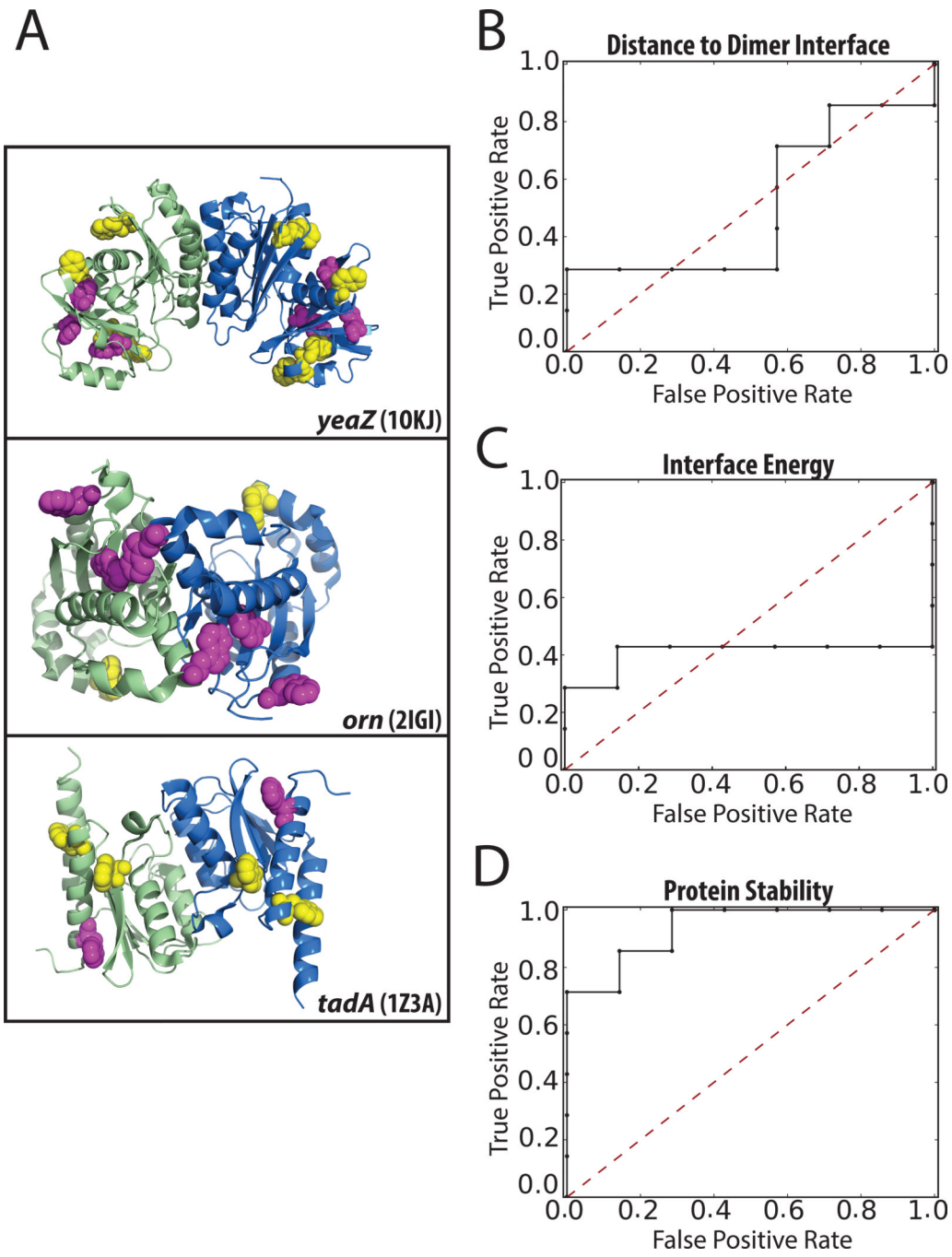

Figure 2. Structural analysis of mutations affecting dimerization

(A) Distribution of W→G mutation sites over the three homodimeric proteins used in the cI assay. Mutation sites that led to loss of repression are shown in *magenta*, the other mutation sites are shown in *yellow*. The dimer subunits are colored *green* and *blue*, respectively. ***(B)*** A receiver operating characteristic (ROC) plot for predicting whether a given mutation will lead to loss of repression in the cI assay, using the distance from the mutation site to the dimer interface as the predictor. The area under the curve is 0.51, indicating that this method performs about as well as making predictions purely at random (the *red dashed line* in each ROC plot corresponds to a random predictor). ***(C)*** An analogous ROC plot generated by

using the difference in interface energy of comparative models to predict whether a given mutation will lead to loss of repression in the cI assay. The area under the curve is 0.41, indicating that this method is not predictive of the data. ***(D)*** An analogous ROC plot generated by using the estimated stability difference from the same set of comparative models. The area under the curve is 0.94, indicating that this method performs much better than a random predictor; the difference from a random predictor is statistically significant ($p < 0.004$). The identification of stability difference as a successful predictor for loss of function suggests that, at least in this experiment, changes in protein stability may underlie inactivation/reactivation.

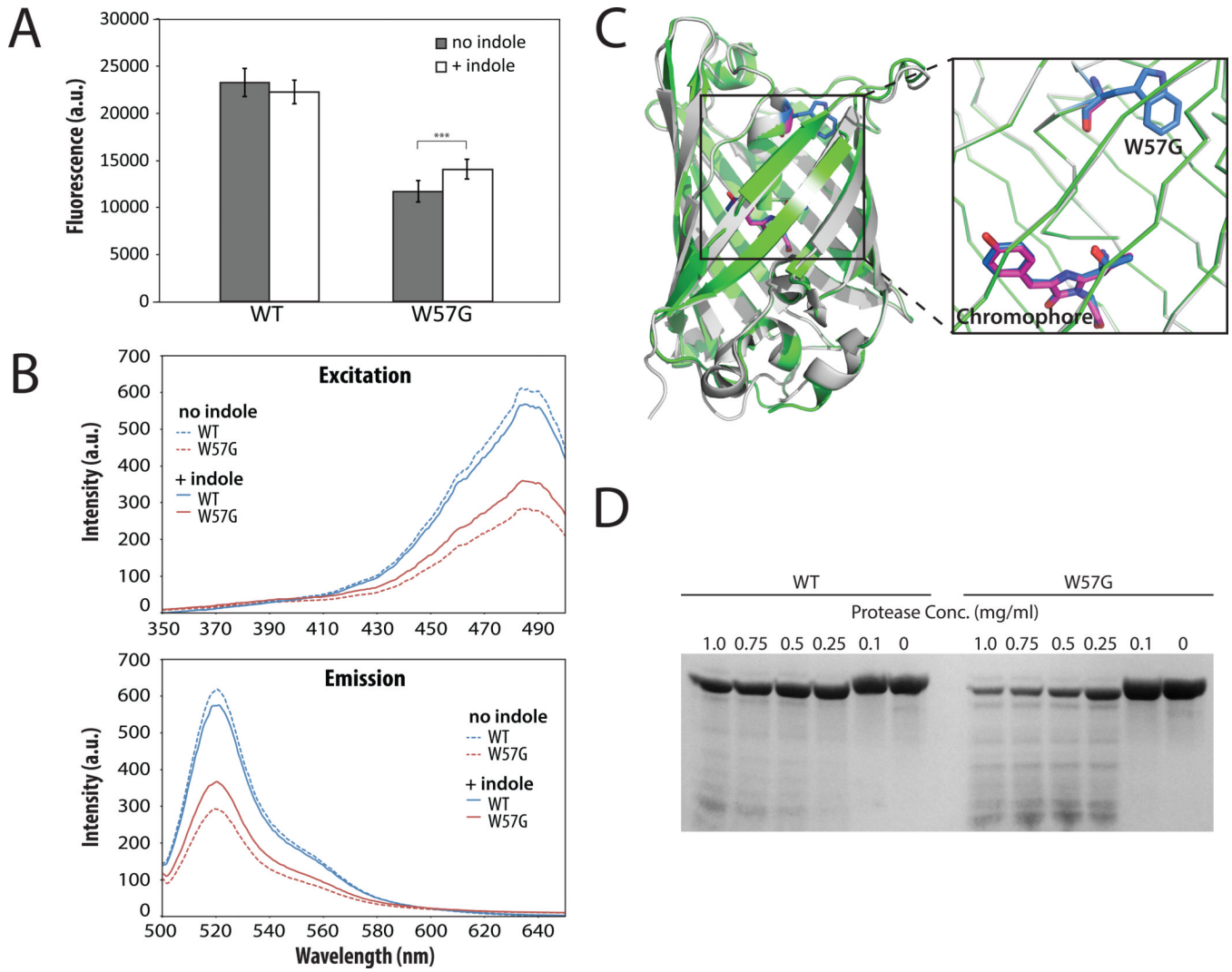

Figure 3. Mechanism of inactivation and rescue in +5 GFP

(A) Fluorescence intensity of +5 GFP constructs, with excitation at 485 nm and emission at 528 nm. The indole concentration was 1 mM. Error bars indicate the standard error of the mean from 10 replicate measurements. *** statistically significant difference at $p < 0.001$. ***(B)*** Excitation and emission spectra of +5 GFP constructs. The indole concentration was 1 mM. ***(C)*** Crystal structure of W57G +5 GFP refined to 1.6 Å resolution (*green* and *magenta*), superposed with wild-type superfolder GFP (*gray* and *blue*). ***(D)*** Coomassie-stained SDS-PAGE gel showing products of pulse proteolysis reactions. Incubation with subtilisin led to more extensive degradation of W57G +5 GFP (*third lane*) than of wild-type +5 GFP (*fourth lane*). Figures S3–S9 include analogous plots showing fluorescence properties of +5 GFP W57A, effect of indole concentration on +5 GFP W57G fluorescence, crystallographic F_o-F_c omit maps, the complete (uncropped) gel from the pulse proteolysis experiment, and a control experiment demonstrating that addition of the protease inhibitor PMSF prevented W57G +5 GFP degradation, while DMSO vehicle alone did not. (Supporting Information).

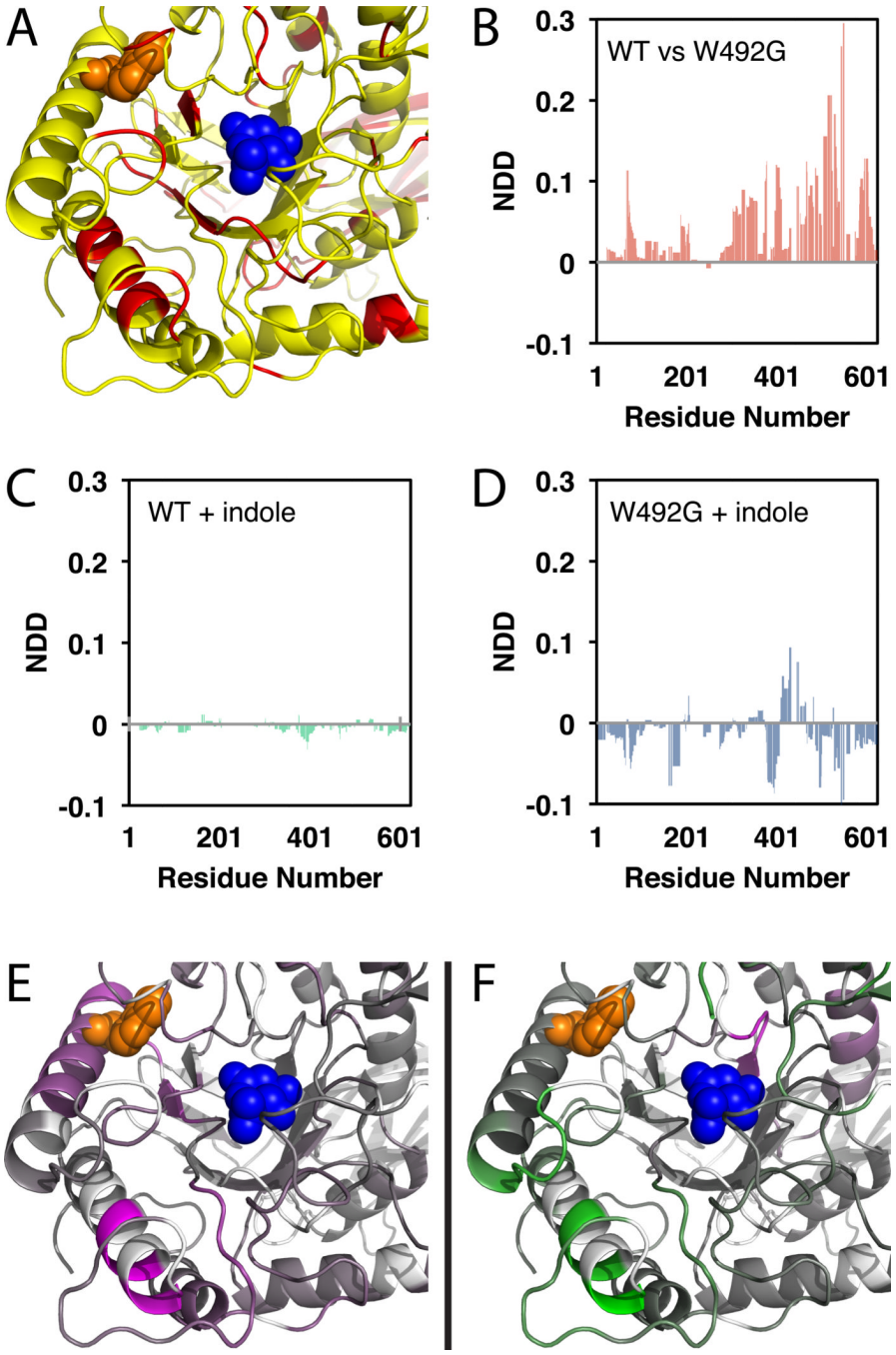

Figure 4. Mechanism of inactivation and rescue in β-glucuronidase as revealed by hydrogen–deuterium exchange analysis

(A) Peptic peptides provide thorough coverage of the β-gluc active site. Residues with an exchangeable backbone amide hydrogen that were not covered by at least one peptide are indicated in *red*; the remainder of the protein is shown in *yellow*. The locations of Trp492 (*orange*) and a substrate analog (*blue*) are shown in spheres. ***(B)*** Comparison of deuterium uptake ("normalized deuterium difference", *NDD*) between wild-type β-gluc and W492G; positive values indicate enhanced deuterium uptake in the mutant. ***(C)*** Effect of adding 5 mM indole to wild-type β-gluc. ***(D)*** Effect of adding 5 mM indole to β-gluc W492G; negative values indicate increased protection from deuterium uptake upon addition of

indole. ***(E)*** Mapping the mutant versus wild-type *NDD* to the β-gluc protein structure reveals a spatial localization of residues that undergo enhanced deuterium uptake in β-gluc W492G relative to the wild-type. The color of each residue reflects the normalized deuterium difference between the mutant and wild-type, using a gradient from *purple* (most enhanced deuterium uptake in the mutant, relative to the wild-type) to *green* (most protected in the mutant, relative to the wild-type). ***(F)*** Mapping the absence versus presence of indole *NDD* to the β-gluc protein structure reveals the pattern of changes that occur upon addition of indole. Each residue is colored using a gradient from *purple* (most enhanced deuterium uptake upon addition of indole) to *green* (most protected upon addition of indole). Figure S13 shows a comparison of deuterium uptake between wild-type β-gluc and W492G with 5 mM indole present in both; the mutant still exhibits greater deuterium uptake, suggesting incomplete rescue at this indole concentration. (Supporting Information).

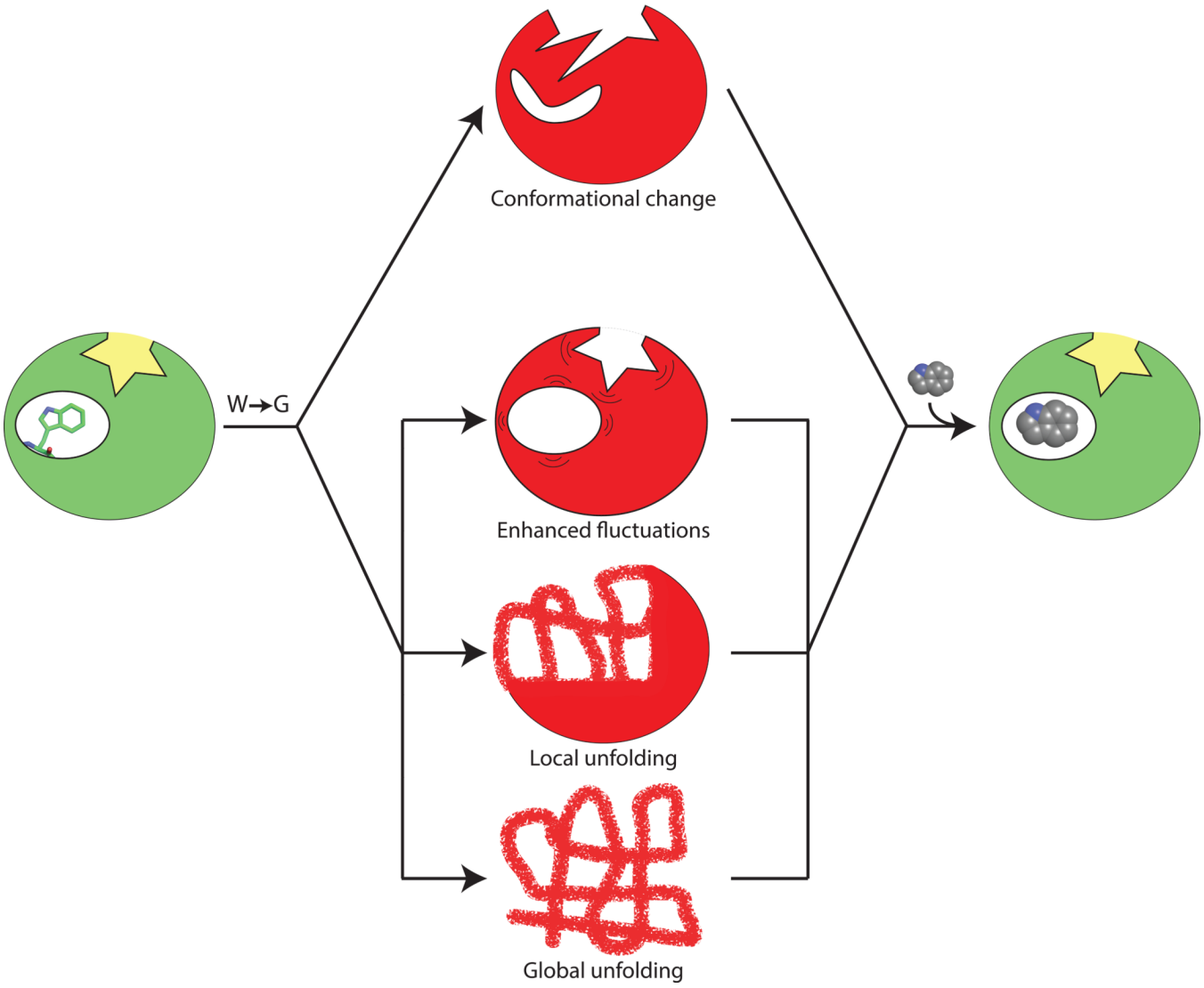

Figure 5. United model of inactivation and rescue

A protein may respond to a cavity-forming W→G mutation by undergoing a discrete conformational change, as seen in our previous study[8], or through stability-mediated mechanisms, as described here. Addition of indole re-activates the protein irrespective of the underlying cause for loss of function.