

# Robustness and evolvability in the functional anatomy of a PER-ARNT-SIM (PAS) domain

Andrew F. Philip<sup>a</sup>, Masato Kumauchi<sup>b</sup>, and Wouter D. Hoff<sup>b,1</sup>

<sup>a</sup>Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637; and <sup>b</sup>Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078

Edited\* by Robert H. Austin, Princeton University, Princeton, NJ, and approved August 6, 2010 (received for review April 9, 2010)

The robustness of proteins against point mutations implies that only a small subset of residues determines functional properties. We test this prediction using photoactive yellow protein (PYP), a 125-residue prototype of the PER-ARNT-SIM (PAS) domain superfamily of signaling proteins. PAS domains are defined by a small number of conserved residues of unknown function. We report high-throughput biophysical measurements on a complete Ala scan set of purified PYP mutants. The dataset of 1,193 values on active site properties, functional kinetics, stability, and production level reveals that 124 mutants retain the characteristic photocycle of PYP, but that the majority of substitutions significantly alter functional properties. Only 35% of substitutions that strongly affect function are located at the active site. Unexpectedly, most PAS-conserved residues are required for maintaining protein production. PAS domain activation often involves conformational changes in  $\alpha$ -helices linked to the PAS core. However, the mechanism of transmission and kinetic regulation of allosteric structural changes from the PAS domain to these helices is not clear. The Ala scan data reveal interactions governing allosteric switching in PYP. The photocycle kinetics is significantly altered by substitutions at 58 positions and spans a 3,000-fold range. Nine residues that dock the N-terminal  $\alpha$ -helices of PYP to its PAS core regulate signaling kinetics. Ile39 and Asn43 are identified as part of a mechanism for regulating allosteric switching that is conserved among PAS domains. These results show that PYP combines robustness with a high degree of evolvability and imply production level as an important factor in protein evolution.

allosteric signal transmission | molecular evolution | protein structure-function relationship

Two central challenges in understanding proteins are to predict (i) the three-dimensional structure adopted by a protein with a given amino acid sequence (the folding problem), and (ii) the functional properties of a protein with a given three-dimensional structure (the “function problem”). The vastness of amino acid sequence space presents a formidable challenge to addressing these problems. One approach to these questions is evolutionary analysis of sequence and structure databases. This strategy has led to the important insight that only a limited number of protein folds exists. In addition, it has revealed superfamilies of protein domains defined by a small number of conserved residues (1, 2). These residues could either constitute the active site or be required for folding. The role of conserved residues in folding is debated (3, 4). Large-scale database analysis has revealed that down to 30% sequence identity active site chemistry in protein superfamilies is maintained (5). However, the level of sequence conservation in most superfamilies is much lower than this. A related discovery is that enzyme superfamilies can be mechanistically diverse and catalyze different chemical reactions (6, 7). Therefore, the few residues that are conserved in protein superfamilies do not appear to be part of a conserved active site, which leaves the functional role of the small number of conserved residues in protein superfamilies unexplained.

A second approach to exploring amino acid sequence space utilizes mutagenesis. Such studies typically evaluate the effects

of mutations through in vivo processes that are easily detected, such as plaque formation (8) and antibiotics resistance (9). However, these processes can be affected by many different properties of the protein, including its folding and functional activity. Here we resolve this complication by dissecting the effects of mutations on specific in vitro biochemical properties of a library of purified mutant proteins. Extensive mutagenesis studies have revealed a high degree of robustness in proteins, where only ~15% of point mutations cause loss of function (8–10). Influential ideas that guide thinking about protein structure–function relationships are that structurally critical residues are often buried and part of the hydrophobic core, whereas functionally critical residues are largely confined to the protein active site. The latter view is challenged by the results reported here.

Robustness against point mutations is a general feature of proteins and prevents frequent catastrophic loss of function during protein evolution. It is often assumed that this robustness implies that most substitutions do not significantly alter the functional properties of a protein. This consideration is in line with Kimura’s neutral theory of evolution, in which most mutations do not alter the properties of the organism (11). We test this proposal using high-throughput methods to examine functional properties of all mutants from a complete Ala mutagenesis scan of a PER-ARNT-SIM (PAS) domain protein.

The PAS domain is a ubiquitous protein module with a characteristic three-dimensional fold (12, 13). The term PAS is derived from the first letter of the name of the three founding members of the superfamily: the *Drosophila* period protein, the aryl hydrocarbon receptor nuclear translocator (ARNT), and the *Drosophila* single-minded protein. Over 20,000 different PAS domains have been identified in a wide variety of signaling proteins in organisms ranging from bacteria to humans. PAS domains consist of ~100 residues, only nine of which are well conserved (12) (Fig. S14). PAS domains can bind a variety of cofactors (12) and exhibit a wide range of active site chemistries. Cofactors at the active sites of PAS domains including *p*-coumaric acid (*p*CA) in photoactive yellow protein (PYP), heme in the oxygen sensor FixL, and flavin in LOV photoreceptors, but some function in the absence of cofactors. In addition, PAS domains exhibit a wide range of active site chemistries, exemplified by *p*CA isomerization and protonation in PYP (14) and transient formation of a Cys-flavin adduct in LOV (15). Thus, PAS domains do not share a conserved active site, and the molecular role played by the residues that are conserved in this superfamily is not clear. The diversity of the PAS domain superfamily provides an opportunity to examine how proteins evolve to perform a wide range of functions. What is the percentage of mutations that alter functional properties? How do these mutations relate to the

Author contributions: W.D.H. designed research; A.F.P. and M.K. performed research; A.F.P., M.K., and W.D.H. analyzed data; and W.D.H. wrote the paper.

The authors declare no conflict of interest.

\*This Direct Submission article had a prearranged editor.

<sup>1</sup>To whom correspondence should be addressed. E-mail: wouter.hoff@okstate.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1004823107/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1004823107/-DCSupplemental).

three-dimensional structure of the protein? These questions are also relevant to the debate on the extent to which amino acid substitutions are selectively neutral or adaptive (11). Here we quantify these key aspects of protein evolvability in a photosensory PAS domain.

PYP is a photoreceptor for negative phototaxis in the halophilic Proteobacterium *Halorhodospira halophila* (16–18). The protein is regarded as a prototype PAS domain (19) and was the first PAS domain to have its high-resolution structure determined (20). PYP exhibits a light-triggered photocycle (17) based on its *pCA* chromophore (21, 22). The *pCA* is enclosed by the protein (20) and is covalently attached to Cys69 (23). The function of PYP involves strong shifts in the  $pK_a$  and absorbance maximum of the *pCA* compared to its properties in solution (24). PYP consists of 125 residues divided into two regions: a typical PAS domain fold (13) with a central antiparallel six-stranded  $\beta$ -sheet flanked by three  $\alpha$ -helices (the PAS domain core), and two N-terminal  $\alpha$ -helices (residues 1–27, the N-terminal region). The N-terminal helices pack against the central  $\beta$ -sheet, forming a second, small hydrophobic core. Structure–function relationships in PYP have been studied by site-directed mutagenesis, revealing a number of residues that are critical to its function (24). Photoexcitation causes the conversion of the initial pG dark state of PYP into the long-lived pB state. The pB state has a blue-shifted absorbance spectrum and is partially unfolded (25–29). Its formation involves dissociation of the N-terminal region from the core of PYP (30, 31). The pB state is believed to be the signaling state of PYP, and spontaneously returns to the pG state in a few hundred milliseconds.

## Results

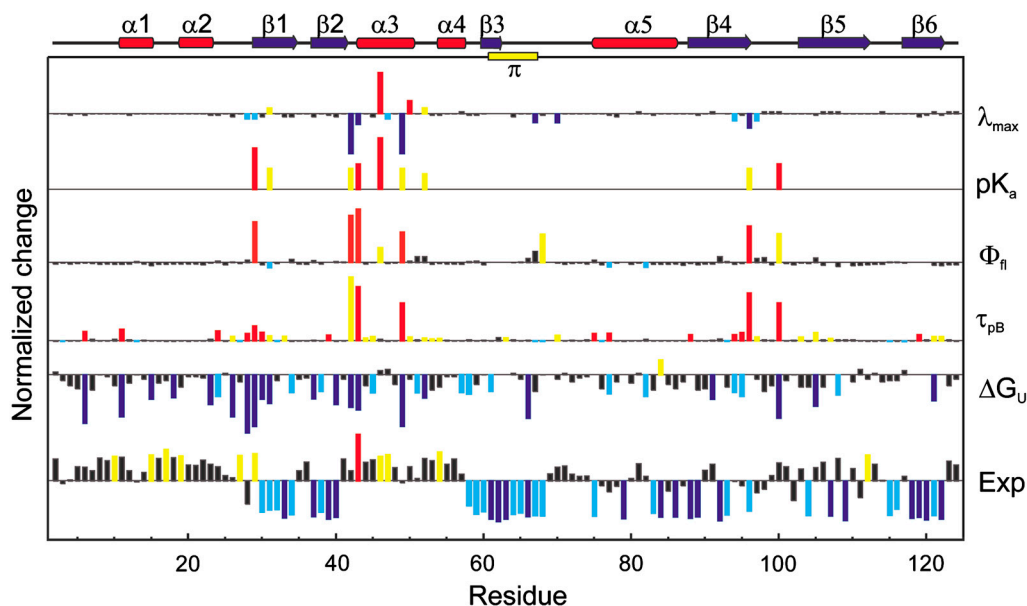
**Robustness in PYP Determined by 96-Well Biophysics.** We have developed 96-well-based methods for the heterologous expression, purification, and spectroscopic characterization of PYP mutants (32). Here, we use this microscale high-throughput approach to study a complete Ala scan library of PYP mutants. The residue at each position was substituted with Ala; positions containing Ala in WT PYP were substituted with Gly. Each of the 125 resulting point mutants was quantitatively examined with respect to six properties: (i) visible absorbance maximum ( $\lambda_{\max}$ ), (ii) *pCA*  $pK_a$ ,

(iii) *pCA* fluorescence quantum yield ( $\Phi_{fl}$ ), (iv) lifetime of the pB state ( $\tau_{pB}$ ), (v) thermodynamic stability against unfolding ( $\Delta G_U$ ), and (vi) protein production level (see *SI Materials and Methods*). The total dataset reported here consists of 1,193 values (Table S1). These microscale high-throughput data are in very good agreement with published results (Table S2). In addition, we examined selected mutants by mass spectrometry. In all cases, the measured mass corresponded to that predicted for the mutant, and no cases of cross-contamination of different samples during growth in 96-well plates were identified.

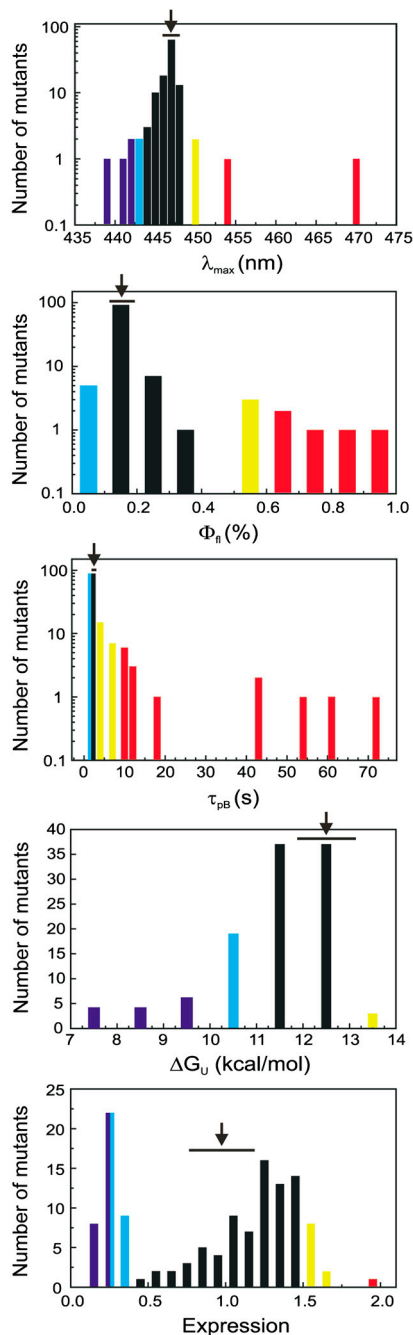
We examined the in vitro functional activity of the 125 PYP mutants (Fig. 1). All but one of the mutants retain a photocycle with a pB intermediate. The only mutation causing loss of function is C69A, in which the chromophore is no longer covalently attached to the protein (23). In addition, all mutants (see *Materials and Methods*) retain a significant  $\Delta G_U$ . This result demonstrates a very high degree of robustness of the structure and function of PYP against point mutations. The level of robustness of PYP against point mutations is approximately 20-fold higher for that reported for other proteins (8–10), likely because in previous studies functional proteins with greatly reduced kinetics or production level would be scored as nonfunctional based on in vivo essays.

**Quantifying Evolvability in PYP.** The Ala scan dataset allows us to quantify the evolvability of PYP with respect to the number of residues that alter functional properties and the range of values that these properties adopt. Although substitutions at only 0.8% of all positions cause the loss of photochemical activity, ~60% significantly alter one or more of the six properties studied here. Active site properties are affected by ~15% of the residues in PYP. Mutations at 58 residues affect  $\tau_{pB}$  (Fig. S2) by more than 50% compared to WT PYP. Substitutions at 53 positions alter the  $\Delta G_U$  of PYP by more than 1 kcal/mol. Finally, 51 substitutions change the production level of PYP by more than 50%.

The values of the functional properties in the PYP mutant set varies over a wide range (Fig. 2 and Table 1). Visible  $\lambda_{\max}$  values range from 439 to 468 nm, the  $pK_a$  of the *pCA* ranges from 3 to 8, and *pCA*  $\Phi_{fl}$  values vary 13-fold (from 0.07% to 0.94%). The value of  $\tau_{pB}$  is altered more than 3,000-fold (from 0.5 s to 30 min).



**Fig. 1.** Mapping changes in functional and structural properties caused by point mutations on the amino acid sequence of PYP. The effect of Ala substitutions on visible absorbance maximum  $\lambda_{\max}$ , *pCA*  $pK_a$ , *pCA* fluorescence quantum yield  $\Phi_{fl}$ , pB lifetime  $\tau_{pB}$ , thermodynamic stability  $\Delta G_U$ , and protein production level are shown as the value of [(value of mutant – value of WT)/value of WT]. The secondary structure of PYP is also shown, with  $\alpha$ -helices in red,  $\beta$ -strands in blue, and the  $\pi$ -helix in yellow. Positions at which substitutions cause large increases in values are indicated in red, small increases in yellow, large decreases in dark blue, and small decreases in light blue. The same color scheme is used in Figs. 2 and 3 to facilitate visual inspection of the dataset.



**Fig. 2.** Range of functional properties observed in the Ala scan of PYP. The arrows indicate the values of WT PYP, whereas the horizontal bar indicates the error in the measurement. Positions for which substitutions cause large increases in values are indicated in red, small increases in yellow, large decreases in dark blue, and small decreases in light blue.

**Table 1. Quantifying the evolvability of PYP**

	$\lambda_{\max}$	pK <sub>a</sub>	$\Phi_H$	$\tau_{pB}$	$\Delta G_U$	Production
% affected mutants*	20%	7%	23%	52%	51%	74%
% strongly affected mutants*	12% <sup>†</sup>	7% <sup>‡</sup>	9% <sup>§</sup>	20% <sup>¶</sup>	28% <sup>  </sup>	41% <sup>§</sup>
Range of values observed	439–470 nm	3–8	0.07–0.94%	0.5 s–30 min	7.2–13.9 kcal/mol	13–200%
Value of WT PYP	447 nm	3.0	0.14	0.9 s	12.5 kcal/mol	100%

\*From a total of 125 substituted positions.

<sup>†</sup>Altered by at least 3 nm compared to WT PYP.

<sup>‡</sup>Altered by at least 1.5 pH units.

<sup>§</sup>Altered by at least ±50%.

<sup>¶</sup>Altered by at least a factor 5.

<sup>||</sup>Altered by at least 2 kcal/mol.

$\Delta G_U$  values from 7.2 to 13.9 kcal/mol were observed, whereas protein production levels varied from a factor 8 lower to a factor 2 higher than WT PYP. Thus, although PYP displays strong robustness against mutations, it simultaneously exhibits a high degree of evolvability: A typical mutation likely does not abolish function, but will alter the functional properties of the protein. This result indicates that modification of functional properties of a protein during evolution can readily occur.

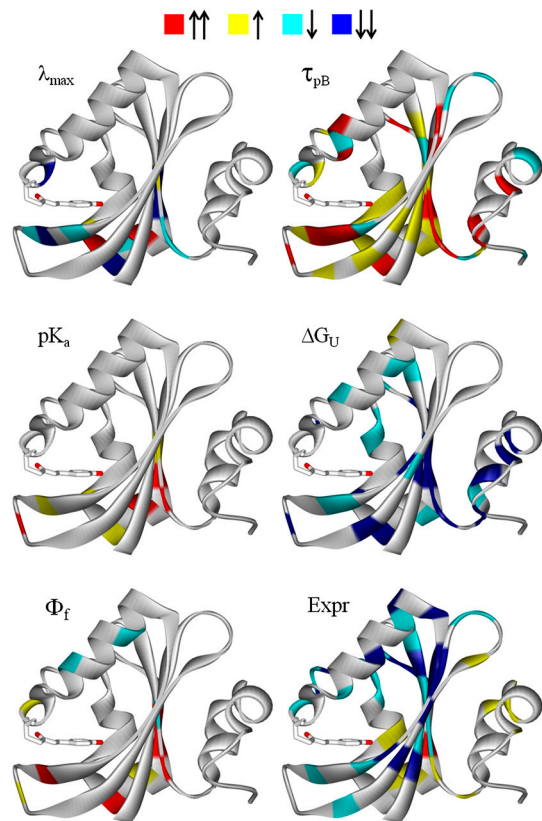
The properties measured here would be expected to be under selective pressure (see *SI Text*). Thus, the finding that substitutions at the majority of the residues in PYP (60%) cause readily detectable effects in these properties argues against the prevalent view (11) that the majority of substitutions are described by Kimura's neutral theory of evolution. The percentage of selectively nearly neutral substitutions in *Drosophila* was recently reported to be 46% (33), very similar to the 40% of substitutions in PYP that do not significantly alter the properties measured here. The results reported here imply that the majority of point mutations in proteins cause changes in functional properties expected to be under selective pressure.

**Protein Production Level as a Key Factor in Amino Acid Sequence Conservation.** Unexpectedly, substitutions at 39 positions decrease PYP production level by more than 60%. Protein production levels vary by 16-fold in the PYP mutants reported here. SDS-PAGE analysis (Fig. S3) of the soluble and insoluble fractions of the *Escherichia coli* cells expressing these PYP derivatives indicated the insolubility of some mutants and proteolytic degradation (little PYP detected) of others, similar to published results on bacteriophage  $\lambda$  Cro (34).

We examined the possibility that the observed reduced PYP production levels are due to mutations in the plasmid or due to the 96-well format of protein expression and purification. We selected the 10 mutants with very low PYP production levels (D34A, G59A, K64A, A67G, P68A, F75A, E93A, K104A, G115A, and F121A) and recloned the *pyp* genes encoding these mutants. After confirmation of the DNA sequence of the inserts, these 10 mutants and WT PYP were expressed in individual 15 mL cultures and purified by batch chromatography using Ni-NTA agarose. Expression levels were evaluated based on the pCA absorbance band of the purified proteins at neutral pH. The results confirmed that the 10 mutants indeed have strongly reduced protein production levels (Fig. S3E).

The data reported here on substitutions that alter the production level of PYP concern the artificial heterologous overexpression of the PYP from *H. halophila* in *E. coli*. However, a number of observations provide strong support that this system yields relevant information on structural features that affect protein production. First, the residues that significantly reduce production are structurally cohesive: They are present as a cluster in the core of PYP (see below). Secondly, 67% and 61% of the residues in PYP that are highly conserved in the PYP family and the PAS domain superfamily significantly reduce the protein overproduction level, respectively (see below). For comparison, less than 30% of all residues in PYP result in a similar level





**Fig. 3.** Mapping the degree to which point mutations alter functional and structural properties on the crystal structure of PYP. The orientation of PYP was chosen to illustrate the packing of the two N-terminal helices to the central  $\beta$ -sheet of PYP. The *pCA* chromophore and Cys69 are shown in stick representation. Substitutions that significantly alter the six indicated functional properties are color coded according to the effect of substitutions at these sites, with large increases in values in red, small increases in yellow, large decreases in dark blue, and small decreases in light blue.

of reduced production. This link of production level to conserved residues supports the physiological relevance of the production levels reported here.

The existence of complex regulatory mechanisms in all organisms to control the expression level of proteins implies that this is a property under strong selective pressure. The strong effect of 31% of the residues in PYP on its production suggests that constraints imposed by maintaining an adequate production level are an important factor in protein evolution.

**Functional Anatomy of PYP.** The high-resolution crystal structure of PYP (20) allows a detailed structural analysis of the Ala scan results (Fig. 3 and Fig. S4). Residues at the active site of a protein often exhibit unusual properties, such as strongly shifted  $pK_a$  values, that are functionally important. How do proteins achieve such tuning of active site residues? Our data identify all residues that are central to functional tuning in the active site of PYP.

Substitution of residues F62, V66, P68, T95, Y98, which are within 5 Å of the *pCA*, does not significantly alter the active site properties of PYP. The residues with the largest effects on the properties of the *pCA* are Gly29, Ile31, Tyr42, Asn43, Glu46, Ile49, Arg52, and Phe96. Of the 20 residues (Table S3) that significantly affect one or more property of the *pCA*, 15 are buried or partially buried and 5 are exposed. Of these side chains, 12 are within 5 Å of the *pCA*, 5 are between 5 and 10 Å, and 3 are more than 10 Å removed from the *pCA*. Thus, most residues that affect the *pCA* are in its immediate vicinity and are buried, matching the traditional view of a protein active site. The interactions

between these residues and the *pCA* include hydrogen bonding, charge–charge interactions, and aromatic ring–ring interactions (Table S3). However, substitutions at four positions far removed from the *pCA* alter its properties by long-range effects: F28A, G77A, G82A, and Y94A. The F28A mutation may shift the *pCA*  $\lambda_{max}$  through altered docking of the N-terminal  $\alpha$ -helices against the  $\beta$ -sheet. Gly77 and Gly82 are located in  $\alpha$ -helix 5; Ala substitution at these positions may decrease *pCA*  $\Phi_f$  and increase  $\tau_{pB}$  by reducing the structural flexibility of this helix. The Y94A mutation removes the side-chain hydrogen bond from residue 94 to the backbone carbonyl oxygen of Cys69, to which the *pCA* is attached. In summary, the interactions that result in active site tuning are complex and heterogeneous.

The lifetime of the pB state is important for its signaling function (see *SI Text*). The 58 substitutions (45% of PYP) that significantly affect  $\tau_{pB}$  are distributed throughout the protein (Fig. 3). This dispersed distribution is in line with the partially unfolded nature of the pB state (25–29): Its decay involves refolding of many regions of the protein, and can thus be affected by mutations at many different locations. The statistically significant correlation of 0.59 between  $\Delta G_U$  and  $\tau_{pB}$  in the Ala scan dataset (Table S4) supports a link (25) between protein folding and pB decay. Analysis of the location of the residues that affect  $\tau_{pB}$  reveals the importance of interactions between the N-terminal region of PYP and its central  $\beta$ -sheet (Table S3). First, the side-chain hydrogen bonds of Asp24 and Asn43 form functionally important connections between the N-terminal domain and the photoactive core of PYP. Second, mutations in 7 of the 10 residues that form the secondary hydrophobic core of PYP greatly increase  $\tau_{pB}$ . Interestingly,  $\tau_{pB}$  is also increased in deletion mutants of PYP lacking the N-terminal region (30, 35). The Ala scan results indicate that both hydrogen bonding and hydrophobic interactions between the N-terminal region and the  $\beta$ -sheet in PYP regulate pB lifetime.

The Ala scan data also yield insights into the structural basis of the  $\Delta G_U$  of PYP. Of the 23 mutations that reduce  $\Delta G_U$  by  $\geq 2$  kcal/mol, the majority is buried (68%) or partially buried (23%), in line with results on other proteins (8, 9). Of these residues, 57% are part of the two hydrophobic cores of PYP. A reduction in  $\Delta G_U$  is caused by substitution of each residue that is part of the minor hydrophobic core, but only 4 of the 20 residues at the major hydrophobic core. We propose that the thermodynamic fragility of the minor hydrophobic core is functionally important by priming the stimulus-induced release of the N-terminal region. NMR measurements of PYP have revealed that the two  $\alpha$ -helices in the N-terminal region are structurally dynamic (36), suggesting that the N-terminal region is not strongly associated with the PAS domain core of PYP. This issue is relevant to signaling by PYP, because dissociation of the N-terminal region is a key event in pB formation (30, 31). Substitutions in the secondary hydrophobic core formed by the packing of the N-terminal region against the central  $\beta$ -sheet not only destabilize PYP by  $\sim 4.0$  kcal/mol, but also increase  $\tau_{pB}$ . This result shows that although the N-terminal region is dynamic, it is thermodynamically coupled to the PAS core, and thus influences both  $\tau_{pB}$  and  $\Delta G_U$ .

How do proteins achieve their remarkable robustness of structure and function against point mutations? Regarding protein stability, the prevalent view is that the strongest effects of single substitutions usually are at most a modest loss of  $\Delta G_U$ , whereas an influential notion on protein function is that only substitutions of the few residues involved in active site chemistry strongly alter functional properties. The PYP Ala scan dataset confirms the above view on protein stability, with substitutions at 28% of the positions causing a significant reduction in  $\Delta G_U$  and losses in stability of up to 5.3 kcal/mol. However, 40% of the residues that alter PYP active site functional properties are not located at the active site (Table S3), whereas only 12% of substitutions that

**Table 2. Summary of effects of substitution of residues in PYP that are highly conserved in the PAS domain superfamily**

Residue	Dist. to			$pK_a$	$\tau_{pB}/\%bl^{\dagger}$	$\Delta G_U^{\ddagger}$	Production level
	$pCA^*$	$\lambda_{max}$	$\Phi_{fl}$				
WT	NA	447	0.14	3	0.9/24	12.5	1.0
Asp34	III	445	0.18	ND	0.9/15	<u>10.9</u>	<u>0.27<sup>s</sup></u>
Gly37	III	445	0.17	3	2/30	<u>10.3</u>	<u>0.24</u>
Ile39	II	446	ND	3	<u>7/75</u>	ND	<u>0.18</u>
Asn43	II	<u>441</u>	<u>0.94</u>	<u>5.5</u>	<u>59/91</u>	<u>9.3</u>	<u>2.0</u>
Gly51	II	447	0.22	3	2/32	<u>10.8</u>	1.04
Gly59	III	446	0.16	3	1/21	11.6	<u>0.28</u>

NA, not applicable; ND, not determined.

\*II indicates 5 and 10 Å; III indicates more than 10 Å.

<sup>†</sup> $\tau_{pB}$  indicated in seconds.

<sup>‡</sup> $\Delta G_U$  is indicated in kcal/mol.

<sup>s</sup>Underlined values deviate significantly from WT PYP.

cause strong changes in the kinetics of pB decay are at the active site (Table S3). These results show that most residues (65%) with the largest effects on PYP function are not located at the active site.

The results reported here reveal that many point mutations significantly reduce protein production level. The structural factors that affect protein production are of considerable biotechnological interest. The side chains of 26 residues whose substitution significantly reduce expression level are fully or partially buried and form a cluster in the center of PYP, suggesting their contribution to a folding nucleus. Mutations in any of the residues that form the  $\pi$ -helix in PYP (residue 61–67) strongly diminish protein production level. No correlation was observed between production level and  $\Delta G_U$  (Table S4). We propose that protein aggregation of mutants in the highly crowded cytoplasm may impose significant restrictions on changes in amino acid sequence.

## Discussion

The data reported here vividly illustrate that important challenges remain in understanding protein structure–function relationships. First, none of the six PAS-conserved residues in PYP were selected in structure-based mutagenesis studies of PYP (24). However, all of these residues significantly alter the functional properties of PYP (Table 2). Secondly, there are only six positions in PYP at which a substitution simultaneously affects  $\lambda_{max}$ ,  $pK_a$ , and  $\tau_{pB}$ : Gly29, Ile31, Tyr42, Asn43, Glu46, and Ile49 (Fig. 1 and Tables S1 and S3). Of these functionally most critical residues only Tyr42 and Glu46 were identified in site-directed mutagenesis studies based on the crystal structure of PYP (24). The efficient microscale high-throughput measurements used here to quantitatively characterize the properties of libraries of purified mutants offers an effective approach to obtain rich datasets that provide insights into the functional anatomy of proteins.

Future elaboration of the high-throughput approach described here can provide additional functional information. One important example is the use of the curved temperature dependence of the kinetics of pB decay (27) or the use of fluorescent dyes (37, 38) to achieve high-throughput detection of the extent of light-induced conformational changes. More generally, the approach reported here can be applied to any protein for which (i) sufficient amount of protein can be overproduced and purified by Ni-NTA affinity chromatography, and (ii) the properties of the protein can be detected spectroscopically either by an intrinsic chromophore (aromatic side chains or cofactors) or by attached fluorescent probes or chromogenic substrate derivatives.

The results reported here allow an examination of the role of the residues that are highly conserved in the PYP family (24) and in the PAS domain superfamily (12), and thus of possible reasons for their evolutionary conservation. The PYP family of receptors shows considerable sequence divergence, with pairwise sequence

**Table 3. Role of conserved residues in PYP**

	Conserved in PYPs	Conserved in PAS domains
Total	23	6
Active site	8	1
$\tau_{pB}$	14	2
$\Delta G_U$	6	2
Production	14	4

The number of residues affecting the indicated properties is shown.

identities down to 19% (24). Of the 23 residues that are highly conserved in PYP (Fig. S1B and Table S5) only eight are located at the photoactive site of PYP (24). Of these residues, only Tyr42, Glu46, and Cys69 play a classical role as active site residues in that they serve as hydrogen bonding partners (20), proton donor (39), and cofactor attachment site (23). Fifteen highly conserved side chains are far removed from the active site. Substitutions at many of these sites either significantly reduce the  $\Delta G_U$  of PYP (six mutants, average  $\Delta\Delta G$  of  $-3.8$  kcal/mol) and/or strongly reduced its protein production (14 mutants, average reduction in production by 76%). This finding indicates that protein stability and production level are major factors in causing high conservation of residues (Table 3).

The PAS domain superfamily is defined by their characteristic but very weak amino acid sequence similarity. However, the functional role of these conserved residues is not known. The Ala scan results yield insights into the role that these residues play. Of the nine most highly PAS-conserved residues (Fig. S1A), six are present in the PYP from *H. halophila*: Asp34, Gly37, Ile39, Asn43, Gly51, and Gly59. These residues are structurally inconspicuous. None are at the  $pCA$  binding pocket; three are buried and three are solvent exposed. In two mutants (I39A and N43A)  $\tau_{pB}$  is strongly increased, whereas Asn43 also affects the properties of the  $pCA$  (Table S3). Substitutions at four PAS-conserved residues cause a  $\sim 75\%$  reduction in protein production level, while leaving the  $\Delta G_U$  of PYP essentially unchanged. These data yield insights into the role of five highly PAS-conserved residues (Table 2), and provide evidence that the majority of these residues affect protein production (Table 3).

The Ala scan data suggest a conserved role of Ile39 and Asn43 in signaling in PAS domains. The side chain of Ile39 is part of the main hydrophobic core of PYP, whereas the side-chain hydrogen bonds of Asn43 form a helical cap for  $\alpha$ -helix 3 and restrain  $\beta$ -strand 1 at the N terminus of the PAS domain fold. Our in-depth studies of PYP mutants with substitutions at Asn43 indicate that its side-chain hydrogen bonds are important for regulating the kinetics of PAS domain signaling. We propose that these two residues form a conserved mechanism to control the kinetics of the allosteric structural changes upon PAS domain deactivation.

The kinetics of pB decay in PYP are also regulated by the N-terminal region (30, 35). The Ala scan data indicate that a significant fraction of the side chains in PYP is dedicated to tuning the interaction between the N-terminal region of PYP and its PAS domain core, and thus regulate  $\tau_{pB}$  (Table S3). The C-terminal helical extensions of a PAS domain from the plant photoreceptor phototropin (40) and the *Drosophila* clock protein Period (41) and the N-terminal extension of a PAS domain in the fungal photoreceptor Vivid (42) all pack against the PAS core of the protein in a position similar to that of the N-terminal region of PYP. Functionally important conformational changes and partial unfolding occur in all of these extensions (30, 31, 40–42). This trend indicates a common theme in PAS domain signaling: partial unfolding of helical extensions that are attached to either the N or C termini of PAS domains. In the case of the heterodimeric transcription factor hypoxia-inducible factor (HIF), functionally important interactions between the PAS domains in HIF $\alpha$  and ARNT occur at the exposed face of the central  $\beta$ -sheet (43). This observation indicates the possibility that the light-induced release of helical

extensions in many PAS domains expose the central  $\beta$ -sheet for subsequent interactions with downstream signaling partners.

## Materials and Methods

**Construction of PYP Mutants.** The PYP mutants were generated with 125 specific mutagenic primers using Stratagene's QuikChange site-directed mutagenesis kit with the *yyp* gene from *H. halophila* in a pQE-80L plasmid (QIAGEN) as the template, and were confirmed by DNA sequencing.

**Protein Purification.** The growth of the 125 *E. coli* strains overproducing the PYP mutants, their purification, and most of their spectroscopic analysis were performed in 96-well format. The proteins were purified using Ni-NTA agarose as described in (32), with the modifications described in the *SI Text*. Selected purified samples were examined by MALDI-TOF mass spectrometry.

- Bateman A, et al. (2004) The Pfam protein families database. *Nucleic Acids Res* 32: D138–D141.
- Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Natl Acad Sci USA* 95:5857–5864.
- Mirny L, Shakhnovich E (2001) Evolutionary conservation of the folding nucleus. *J Mol Biol* 308:123–129.
- Larson SM, Ruczinski I, Davidson AR, Baker D, Plaxco KW (2002) Residues participating in the protein folding nucleus do not exhibit preferential evolutionary conservation. *J Mol Biol* 316:225–233.
- Todd AE, Orengo CA, Thornton JM (2001) Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol* 307:1113–1143.
- Gerlt JA, Babbitt PC (2001) Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem* 70:209–246.
- Glasner ME, Gerlt JA, Babbitt PC (2006) Evolution of enzyme superfamilies. *Curr Opin Chem Biol* 10:492–497.
- Rennell D, Bouvier SE, Hardy LW, Poteete AR (1991) Systematic mutation of bacteriophage T4 lysozyme. *J Mol Biol* 222:67–87.
- Huang W, Petrosino J, Hirsch M, Shenkin PS, Palzkill T (1996) Amino acid sequence determinants of  $\beta$ -lactamase structure and activity. *J Mol Biol* 258:688–703.
- Wagner A (2005) *Robustness and Evolvability in Living Systems* (Princeton Univ Press, Princeton, NY), pp 63–65.
- Nei M (2005) Selectionism and neutralism in molecular evolution. *Mol Biol Evol* 22:2318–2342.
- Taylor BL, Zhulin IB (1999) PAS domains: Internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* 63:479–506.
- Hefti MH, François KJ, de Vries SC, Dixon R, Vervoort J (2004) The PAS fold. A redefinition of the PAS domain based upon structural prediction. *Eur J Biochem* 271:1198–1208.
- Hellingwerf KJ, Hendriks J, Gensch T (2003) Photoactive yellow protein, a new type of photoreceptor protein: Will this “yellow lab” bring us where we want to go? *J Phys Chem A* 107:1082–1094.
- Christie JM (2007) Phototropin blue-light receptors. *Annu Rev Plant Biol* 58:21–45.
- Meyer TE (1985) Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira halophila*. *Biochim Biophys Acta* 806:175–183.
- Meyer TE, Yakali E, Cusanovich MA, Tollin G (1987) Properties of a water-soluble, yellow protein isolated from a halophilic phototrophic bacterium that has photochemical activity analogous to sensory rhodopsin. *Biochemistry* 26:418–423.
- Sprenger WW, Hoff WD, Armitage JP, Hellingwerf KJ (1993) The Eubacterium *Ectothiorhodospira halophila* is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein. *J Bacteriol* 175:3096–3104.
- Pellequer JL, Wager-Smith KA, Kay SA, Getzoff ED (1998) Photoactive yellow protein: A structural prototype for the three-dimensional fold of the PAS domain superfamily. *Proc Natl Acad Sci USA* 95:5884–5890.
- Borgstahl GEO, Williams DR, Getzoff ED (1995) 1.4 Å structure of photoactive yellow protein, a cytosolic photoreceptor—Unusual fold, active site, and chromophore. *Biochemistry* 34:6278–6287.
- Hoff WD, et al. (1994) Thiol ester-linked *p*-coumaric acid as a new photoactive prosthetic group in a protein with rhodopsin-like photochemistry. *Biochemistry* 33:13959–13962.
- Baca M, et al. (1994) Complete chemical structure of photoactive yellow protein: Novel thioester-linked 4-hydroxycinnamyl chromophore and photocycle chemistry. *Biochemistry* 33:14369–14377.
- van Beeumen JJ, et al. (1993) Primary structure of a photoactive yellow protein from the phototrophic bacterium *Ectothiorhodospira halophila*, with evidence for the mass and the binding site of the chromophore. *Protein Sci* 2:1114–1125.
- Kumauchi M, Hara M, Stalcup P, Xie A, Hoff WD (2008) Identification of six new photoactive yellow proteins: Diversity and structure-function relationships in a bacterial blue light photoreceptor. *Photochem Photobiol* 84:956–969.
- Lee B-C, Pandit A, Croonquist PA, Hoff WD (2001) Folding and signaling share the same pathway in a photoreceptor. *Proc Natl Acad Sci USA* 98:9062–9067.
- Bernard C, et al. (2005) The solution structure of a transient photoreceptor intermediate: Delta 25 photoactive yellow protein. *Structure* 13:953–962.
- van Brederode ME, Hoff WD, van Stokkum IHM, Groot ML, Hellingwerf KJ (1996) Protein folding thermodynamics applied to the photocycle of the photoactive yellow protein. *Biophys J* 71:365–380.
- Hoff WD, et al. (1999) Global conformational changes upon receptor stimulation in photoactive yellow protein. *Biochemistry* 38:1009–1017.
- Zhao JM, et al. (2006) Single-molecule detection of structural changes during Per-Arnt-Sim (PAS) domain activation. *Proc Natl Acad Sci USA* 103:11561–11666.
- van der Horst MA, van Stokkum IH, Crielgaard W, Hellingwerf KJ (2001) The role of the N-terminal domain of photoactive yellow protein in the transient partial unfolding during signalling state formation. *FEBS Lett* 497:26–30.
- Imamoto Y, Kamikubo H, Harigai M, Shimizu N, Kataoka M (2002) Light-induced global conformational change of photoactive yellow protein in solution. *Biochemistry* 41:13595–13601.
- Philip AF, Eisenman KT, Papadantonakis GA, Hoff WD (2008) Functional tuning of photoactive yellow protein by active site residue 46. *Biochemistry* 47:13800–13810.
- Sawyer SA, Parsch J, Zhang Z, Hartl DL (2007) Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. *Proc Natl Acad Sci USA* 104:6504–6510.
- Pakula AA, Young VB, Sauer RT (1986) Bacteriophage  $\lambda$  *cro* mutations: Effects on activity and intracellular degradation. *Proc Natl Acad Sci USA* 83:8829–8833.
- Harigai M, et al. (2001) Amino acids in the N-terminal region regulate the photocycle of photoactive yellow protein. *J Biochem* 130:51–56.
- Düx P, et al. (1998) Solution structure and backbone dynamics of the photoactive yellow protein. *Biochemistry* 37:12689–12699.
- Lee B-C, Croonquist PA, Sosnick TR, Hoff WD (2001) PAS domain receptor photoactive yellow protein is converted to a molten globule state upon activation. *J Biol Chem* 276:20821–20823.
- Hendriks J, et al. (2002) Transient exposure of hydrophobic surface in the photoactive yellow protein monitored with Nile red. *Biophys J* 82:1632–1643.
- Xie A, et al. (2001) Formation of a new buried charge drives a large-amplitude protein quake in photoreceptor activation. *Biochemistry* 40:1510–1517.
- Harper SM, Neil LC, Garner KH (2003) Structural basis of a phototropin light switch. *Science* 301:1541–1544.
- Yildiz O, et al. (2005) Crystal structure and interactions of the PAS repeat region of the *Drosophila* clock protein PERIOD. *Mol Cell* 17:69–82.
- Zoltowski BD, et al. (2007) Conformational switching in the fungal light sensor vivid. *Science* 316:1054–1057.
- Erbel PJA, Card PB, Karakuzu O, Bruick RK, Gardner KH (2003) Structural basis for PAS domain heterodimerization in the basic helix-loop-helix-PAS transcription factor hypoxia-inducible factor. *Proc Natl Acad Sci USA* 100:15504–15509.