A hinge migration mechanism unlocks the evolution of green-to-red photoconversion in GFP-like proteins

Hanseong Kim[1§#], Taisong Zou[2§], Chintan Modi[3], Katerina Dörner[1], Timothy J. Grunkemeyer[1], Liqing Chen[1], Raimund Fromme[1], Mikhail V. Matz[4], S. Banu Ozkan[2], Rebekka M. Wachter[1]

[1]Center for Bioenergy and Photosynthesis, Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA

[2]Center for Biological Physics, Department of Physics, Arizona State University, Tempe, AZ 85287, USA

[3]Institute for Cellular and Molecular Biology, The University of Texas, Austin, TX 78712, USA

[4]Section of Integrative Biology, The University of Texas, Austin, TX 78712, USA

CONTACT INFORMATION: Rebekka M. Wachter, phone 480-965-8188, fax 480-965-2747, email rwachter@asu.edu;

FOOTNOTES: [§]These authors contributed equally to this work. [#]Present address: Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI 48109.

TITLE RUNNING HEAD: Structures of GFP-like ancestral proteins

Summary

In proteins, functional divergence involves mutations that modify structure and dynamics. Here, we provide experimental evidence for an evolutionary mechanism driven solely by long-range dynamic motions without significant backbone adjustments, catalytic group rearrangements or changes in subunit assembly. Crystallographic structures were determined for several reconstructed ancestral proteins belonging to a GFP class frequently employed in super-resolution microscopy. Their chain flexibility was analyzed using molecular dynamics and perturbation response scanning. The green-to-red photoconvertible phenotype appears to have arisen from a common green ancestor by migration of a knob-like anchoring region away from the active site diagonally across the beta-barrel fold. The allosterically coupled mutational sites provide active site conformational mobility via epistasis. We propose that light-induced chromophore twisting is enhanced in a reverse-protonated subpopulation, activating internal acid-base chemistry and backbone cleavage to enlarge the chromophore. Dynamics-driven hinge migration may represent a more general platform for the evolution of novel enzyme activities.

Highlights

1. The structures of green and green-to-red photoconvertible ancestors are very similar.

2. In a β-barrel, the locations of fold-anchoring and breathing regions can be reversed.

3. Residue replacements cause increased stiffness at a remote region by allosteric means.

4. A softer active site supports concerted carbon acid deprotonation and β-elimination.

Keywords: Green fluorescent protein; red fluorescent protein; super-resolution microscopy; Kaede; photoconversion; photochromism; photoinduced backbone cleavage; ancestral gene reconstruction; molecular evolution; dynamics-driven protein evolution; global motions; chain flexibility; catalytic base activation; reverse protonation; principle of kinetic equivalence; allosteric communication; epistasis;

Abbreviations: FP, fluorescent protein; GFP, green fluorescent protein; FP, fluorescent protein; TLS, translation, libration, screw rotation; G/R photoconvertible, green-to-red photoconvertible; ENM, elastic network model; PRS, perturbation response scanning; MD, molecular dynamics; *dfi*, dynamic flexibility index; RMSF, root mean square fluctuation; RMSD, root mean square deviation.

Proteins possess unique structure-encoded dynamics that underlie their biological functions (Bahar et al., 2010). Thermal motions range from atomic fluctuations and side chain rotations to collective movements propagated through networks of residues. The modern view of protein structure includes the notion of an ensemble of conformations in equilibrium with each other, and dynamics is considered an important tuning parameter to create new function or to adapt to new environments (Tokuriki and Tawfik, 2009; Glembo et al., 2012; Bhabha et al., 2013; Gerek et al., 2013). Ample precedent exists for the use of dynamics to modulate the promiscuity of substrate binding or the rate of product release (Ortlund et al., 2007). However, a more comprehensive understanding of phenotypic change requires insight into adjustments in chain flexibility while the backbone conformation is maintained, and insight into the mechanism by which the distribution of stiff and loose regions of the protein scaffold affects the chemical steps of catalysis.

As a model system, we have chosen relatives of the green fluorescent protein (GFP) that belong to the order *Scleractinia*, suborder *Faviina* along the phylogenetic clade D of stony corals. For this branch, evolutionary color diversification from an ancestral green-fluorescent form has been described to some detail (Ugalde et al., 2004; Field et al., 2006; Alieva et al., 2008). Red-emitting proteins have been shown to consist solely of the Kaede-type, a class of GFP-like proteins that undergoes irreversible green-to-red photoconversion upon illumination with UV/blue light. Once the green chromophore is formed from an internal cyclization, oxidation and dehydration reaction involving His62-Tyr63-Gly64 tripeptide (Zhang et al., 2006; Pouwels et al., 2008), light activation incorporates the His62 imidazole group into the chromophore's π-system (Wachter et al., 2010). In all Kaede-type proteins, light-excitation is coupled to a β-elimination step that cleaves an internal main chain bond. The photochemistry leads to an enlargement of the chromophore from a two-ring to a three-ring π-system by incorporation of the His62 side chain, thereby red-shifting absorbance and fluorescence. The

elucidation of the photoconversion chemistry is of intense current interest, because Kaede-type proteins are utilized extensively in super-resolution fluorescence microscopy applications (Moeyaert et al., 2014).

To better understand the evolution of red color in *Faviina*, a series of ancestral proteins have been reconstructed in the laboratory (Ugalde et al., 2004; Field et al., 2006). These range from a common green ancestor (ALL-GFP) to a least evolved ancestor (LEA) that harbors the minimum number of mutations necessary and sufficient to generate red color with high efficiency (Field and Matz, 2010). Recently, we have demonstrated that the LEA quantum yield of photoconversion is 1.5×10^{-3}, in line with that of a series of well-characterized extant proteins (Kim et al., 2014).

To generate LEA from ALL-GFP, a total of 13 historical mutations were introduced (Field and Matz, 2010), six with internal side chains near the chromophore (A60V, Q62H, T69A, S105N, Y116N, V157I), one with a side chain in the antiparallel subunit interface (T104R), two with solvent-exposed side chains (M154T and D74H), and four that are part of the C-terminal tail or interact with it (R194C, R216H, ΔY217, M218G, where Δ denotes deletion), see Table S1. With the exception of Q62H, which introduces the imidazole building block for the red chromophore, single-site reversions to the ancestral state do not abolish photoconversion (Field and Matz, 2010). However, two substitutions and one deletion were shown to be most critical in promoting red color, T69A, Y116N, and ΔY217, although a detailed rationale for this observation has not been provided yet (Field and Matz, 2010).

Here, we present several high-resolution X-ray structures of common green ancestors and partially evolved variants thought to recapitulate evolution, and compare these structures to the recently reported X-ray structure of LEA (Kim et al., 2014). Surprisingly, we do not observe any backbone or catalytic group rearrangements within the 11-stranded β-barrel, nor do we observe any perturbation of the tetrameric assembly. However, a computational comparison of dynamic features using Perturbation

Response Scanning (PRS) (Atilgan and Atilgan, 2009; Atilgan et al., 2010) suggests that a reversal in the directionality of protein breathing motions may have occurred. PRS methods have recently been utilized to investigate allosterism (Gerek and Ozkan, 2011; Gerek et al., 2013), and are closely related to structure perturbation models (SPM) developed to compute the residue-dependent response to local perturbation by means of an elastic network model (ENM) (Zheng et al., 2005). Based on our crystallographic and computational results, we propose that in Kaede-type proteins, a softer active site allows for histidine-mediated proton shuttling, ultimately facilitating a concerted mechanism for carbon acid deprotonation and backbone scission. According to this model, the photochemistry is facilitated by increased active-site dynamics allosterically coupled to enhanced rigidity at a remote region.

Results

Photoconversion-promoting mutations destabilize the protein fold. To better understand the evolution of red color in stony corals, we have expressed, purified and crystallized a series of reconstructed ancestral GFP-like proteins in their green pre-photoconversion state. The evolution of Kaede-type proteins is represented by the ancestral non-photochromic variants ALL-GFP, ALL-GFP-new (redesigned C-terminus) and ALL-Q62H, and the evolved photoconversion-competent variants LEA (Ugalde et al., 2004), LEA-X121 (N116Y reversion mutation), LEA-X72 (A69T reversion), and LEA-X6 (V60A, H74D, T154M and I157V reversions) (Field and Matz, 2010). For corresponding GFP residue numbers, see Table S1 and Fig. S1. Illumination with a UV lamp confirmed that photoconversion efficiency decreased in the order LEA, LEA-X6, LEA-X121 and LEA-X72 (Field and Matz, 2010). To examine thermostability, the intrinsic fluorescence of each variant was monitored as a function of temperature (Fig. 1a). The apparent denaturation mid-point $T_{m\text{-}app}$ was shown to decrease from 98° C for ALL-Q62H to 76°C for LEA, indicating that substantial thermal destabilization

accompanies photochromism (Fig. 1a). Strikingly, the single-site reversion with the most dramatic loss of photochromism, LEA-X72 (A69T), almost completely recovers the ancestral thermostability ($T_{m\text{-}app}$ = 94° C), suggesting a role for interior packing. For all variants, dynamic light scattering (DLS) provided hydrodynamic radii consistent with tetrameric species (Kennis et al., 2013), although LEA appeared more prone to aggregation at low salt and low protein concentrations.

The pK_a values of the green chromophore ranged from 5.2 to 6.4 (Table S2), similar to those reported for Kaede (pK_a = 5.6) and EosFP (pK_a = 5.8) (Ando et al., 2002; Wiedenmann et al., 2004). The only exception was LEA-X72 with a pK_a of 8.4, more basic than Dendra2 (pK_a = 7.1)(Adam et al., 2009) or KikGRX (pK_a = 8.0)(Tsutsui et al., 2009), suggesting that a high pK_a correlates with reduced photoconversion efficiency. For all variants, the green fluorescence quantum yield was measured to be ~80% (Table S3 provides a list of optical properties).

Crystallographic models suggest that catalytic groups and core subunit contacts are highly conserved. Crystallization was carried out in the green state of each variant at least 1.3 pH units above the chromophore pK_a. ALL-GFP (1.6 Å resolution) and ALL-Q62H (1.4 Å) crystallized in space group C2 with four molecules per asymmetric unit (Tables 1, 2). Superimpositions of chains A-D provided root-mean-square deviations (RMSD) of 0.17 - 0.20 Å for α-carbons and 0.48 - 0.67 Å for all atoms. LEA-X6 (2.5 Å) and LEA-X121 (1.7 Å) crystallized in space group I222 with one molecule per asymmetric unit (Tables 1, 2). Crystals obtained for LEA-X72 were of insufficient quality for structure determination, and the structure of LEA (1.5 and 1.8 Å) was reported previously by our group (Kim et al., 2014). Pair-wise superimpositions of each variant onto the ALL-GFP X-ray model provided α-carbon RMSD values ranging from 0.14 to 0.29 Å, suggesting that the main-chain conformations remain unmodified for all variants examined.

For the entire set of reconstructed proteins, the tetrameric assembly generated by crystallographic (I222) or non-crystallographic (C2 or $P2_12_12_1$) symmetry consisted of a dimer of dimers, as for DsRed, Kaede and EosFP (Yarbrough et al., 2001; Ando et al., 2002; Wiedenmann et al., 2004). The core of each molecular contact surface appears highly conserved across all variants. The A - B (and C - D) subunit interface is formed by a nearly perpendicular arrangement of β-barrels and is dominated by hydrophilic contacts, whereas the A – D (and B – C) interface is formed by a nearly antiparallel arrangement of β-barrels and is dominated by van der Waals, hydrogen and ionic interactions.

The coplanar preorganization of the green chromophore with the side chain of residue 62 (Gln in ALL-GFP, otherwise His) was found to be highly conserved across all variants irrespective of photoconversion competency (Fig. S2). The conformation of the catalytic base Glu211 was found to be equally conserved, with the carboxylate in van der Waals contact with the β-carbon of His62 (~3.5 Å). In addition, the active site network of hydrogen bonds between Glu211, Gln38, water and Phe61 was found to be conserved (Fig. 1bc) (Kim et al., 2014). Therefore, these features may be necessary but not sufficient for photochromic behavior.

Repacking of the chromophore's phenolic end may guide light-induced twisting to displace His193. A comparison of the ALL-Q62H and LEA structures indicates that the critical T69A substitution disrupts a hydrogen bond between the Thr hydroxyl and N_ε of Arg66, thereby reorienting the Arg66 side chain to instead H-bond with the chromophore's carbonyl oxygen (Fig. 2a). This movement breaks the ionic bond with Glu211, leaving the carboxylate with only one salt bridge, that to His193 (Fig. 1c). To fill the vacated space, His193 is shifted towards the protein's interior by about 0.5 Å. Intriguingly, the near-by V157I substitution raises the number of van der Waals contacts with the chromophore and Phe173 near the opposite face of the chromophore. Therefore, T69A and V157I may work in concert to direct chromophore twisting towards His193. The emerging model suggests that

light-induced bond rotations around the chromophore's methylenic bridge orient the phenolic group to displace the π-stacked His193 imidazole group.

Reduced packing of the His62 imidazole is combined with optimized imidazolinone anchoring. The Y116N substitution enlarges a solvent-bearing cavity adjacent to the His62-derived imidazole group, thereby accommodating an additional ordered water molecule (Fig. 1c). We propose that this feature allows for His62 rotamer adjustments to shuttle protons from Glu211 to the carbonyl oxygen of Phe61. Moreover, Y116N disrupts a van der Waals contact with Phe68, likely augmenting the T69A packing defect. To compensate, anchoring of the chromophore's backbone-derived imidazolinone ring appears to be optimized by the A60V, T104R and S105N mutations, a set that appears to provide core-stabilizing interaction. A60V serves to increase molecular packing of the helical turn immediately preceding the chromophore, whereas T104R forms an intra-chain salt bridge to Glu90 by placing a larger, charged side chain into the A – B interface. The resulting 0.5 Å backbone shift of residues 104 and 105 towards the interior facilitates an H-bond between S105N and Gly64.

Disordering of the C-terminus releases segments of β-strands lining the chromophore. In the tetrameric assemblies of the ALL variants, the C-terminal clasp residues 216 – 225 exhibit well-defined electron density and contact the adjacent chain in the perpendicular subunit interface. Surprisingly, in LEA and its variants, residues 220 – 225 are found to be crystallographically disordered (Fig. 2bc). Among the set of mutations affecting C-terminal structure (R194C, R216H, ΔY217, M218G), R216H disrupts a hydrogen bond to adjacent-chain Arg216, whereas the deletion of Tyr217 places the remainder of the peptide out of register. In this way, a series of intermolecular interactions are eliminated involving residues 217 – 225 and neighboring chain residues 141 (strand #7), 192, 194, 196 and 198 (strand #10), and 210 and 212 (strand #11). Since these β-strand segments form van der Waals

contacts with the chromophore via Ser142 (strand #7), His193 (strand #10) and Glu211 (strand #11), the release of the C-clasp may increase the deformability of the chromophore-bearing pocket.

Molecular dynamics simulations suggest an increased range of motions for the photoconversion-competent chromophore. Although the crystallographic volumes of the chromophore-bearing cavities of ALL-Q62H and LEA are of similar size (Fig. S3), internal mutations and C-tail disordering may affect dynamic features of the chromophore. To better assess changes in flexibility, we carried out molecular dynamics (MD) simulations of the ALL-Q62H and LEA tetramers. These simulations solely address dynamic motions of the pre-photoconversion (green) state. Upon 100 ns equilibration, protein motion was simulated for 220 ns with the chromophore remaining in its electronic ground state (Fig. S4). We find that the root mean square fluctuations (RMSF) of the four chromophores are significantly higher in LEA than in ALL-Q62H (Fig. 3a), suggesting that chromophore distortions away from planarity are substantially more accessible in the evolved variant.

Perturbation response scanning (PRS) reveals a close relationship between chain dynamics and functional divergence. To analyze global changes in protein flexibility, PRS methods were applied to the tetrameric species of ALL-Q62H and LEA. In this approach, an isotropic perturbation is applied to each residue one-by-one, and for each step, the ensuing displacement of all remaining residues is computed. To incorporate the effects of side chain modifications introduced by mutagenesis, covariance matrices were calculated using all-atom MD simulations. The trajectories extracted from the α-carbon positions were then utilized to calculate the dynamic flexibility index (*dfi*) of each residue. In this approach, side chain dynamics is intrinsically taken into account, as the trajectories represent the motions of all atoms in a particular residue. PRS is similar in approach to SPM, as both methods measure the impact of a perturbation on dynamics. However, in PRS, the perturbation consists of a Brownian kick applied to one residue at a time, whereas in SPM, mutational effects are simulated by

altering spring constants in ENM representations (Zheng et al., 2005). Because the backbone structures of ALL-Q62H and LEA are nearly identical, ENM methods proved insufficient to establish any differences, whereas PRS with equilibrated MD trajectories provided clear evidence of modified flexibility. Accordingly, *dfi* values were determined by first calculating the mean square response fluctuation of each residue upon perturbation elsewhere, then normalizing this value with respect to the average mean square response fluctuation of all residues (Eqn. S4) (Ikeguchi et al., 2005). In general, a high *dfi* value would be associated with a high degree of flexibility, whereas a low *dfi* value would indicate transfer of the perturbation energy to the surroundings.

To examine global chain flexibility, the *dfi* values were rank-ordered, assigned a percentile (Fig. 3b) and mapped onto the protein fold (Fig. 4). Not surprisingly, the subunit interfaces bear low *dfi* values in both variants (blue tones), in line with depressed B-factors, and the exterior surfaces generally bear higher *dfi* values (red tones), in line with elevated B-factors. However, striking differences were observed for two corners of the β-barrel, where high *dfi* values in ALL-Q62H were found to correspond to low *dfi* values in LEA, and vice versa. To identify sites with substantially modified dynamics, the residue-by-residue difference in percent *dfi* (Δ*dfi*) between the two proteins was calculated (Fig. 3b). In this way, 16 sites with elevated *Δdfi* (designated "red") and 13 sites with depressed *Δdfi* (designated "blue") were flagged as the top-scoring 15% (Fig. 3 and 5). Strikingly, these sites were found to be distinctly segregated in space, with red sites clustered in a corner near the bottom of the barrel close to the chain termini, and blue sites clustered in loop regions lining the top of the barrel near the chain interfaces (Fig. 5c). No obvious correlation was observed between *Δdfi* values and differences in normalized crystallographic B-factors (Figure S5). The clear separation of regions of increased and decreased flexibility appears to imply that a rigid knob-like component has shifted diagonally across the β-barrel in response to the historical mutations correlating with photochromism.

Regions with increased dynamics map to an internal peptide and to β-strand segments near the bottom of the barrel. A largely buried, irregularly structured peptide comprising residues 67 – 71 is classified as red, and extends from the 3_{10}-helical turn (65 - 68) attached to the chromophore to the β-barrel cap bearing the N- and C-termini (Fig. 5ab). Likely, increased fluctuations are caused by T69A (Fig. 2a) and augmented by Y116N, which eliminates a contact with Phe68. Motions may be propagated via main-chain H-bonding from Lys70 to the red residues 213 – 215 (strand #11), from there to the red residues 187 – 193 (adjacent strand #10), and from there to red Ser142 (adjacent strand #7). These segments include the two red residues in direct contact with the chromophore, Ser142 (strand #7) and His193 (strand #10). These may play a critical role in photoconversion, as chromophore twisting requires the transient disruption of H-bonding to Ser142 and π-stacking with His193 (Fig. 6). The emerging model suggests that the T69A-based elimination of a methyl and hydroxyl group about 8 Å distant from the chromophore's methylenic bridge has the effect of softening the chromphore-bearing pocket. This feature appears to be augmented by unstructuring of the C-terminal clasp residues 220 – 225 that are not included in the *dfi* analysis.

A knob-like region with decreased dynamics maps to the top ridge of the β-barrel. The 13 blue residues with substantially depressed *Δdfi* values map to five loop regions lining the cap diagonally across the β-barrel from the red segment (Fig. 5ab). The rigidified corner consist of residues 19 - 22, 50, 97 - 98, 125, 127 - 129, 165 and 167, residues that form a cluster near but not in the antiparallel interface, and appear to serve as a hinge region stabilizing the evolved variant. Only three blue residues are involved in subunit contacts, with Lys22 (strand #2) forming an ionic bond with neighboring chain Glu117 (strand #6), and Asn19 forming an H-bond with neighboring-chain Lys178. Blue Pro127 and same-chain Asn128 (loop connecting strands #6 and #7) make van der Waals contact with neighboring-chain Asp150 (loop connecting strands #7 and #8). In combination, these interactions suggest that the

antiparallel interface may serve to fortify the hinge region, possibly to compensate for a weakening of the perpendicular interface induced by C-tail disorder.

Most of the mutational sites do not map to red or blue regions. Strikingly, the mutational sites themselves exhibit mostly average *Δdfi* values, suggesting neither elevated nor depressed motions (Fig. 5cd). The chromophore itself, which includes the Q62H mutation, exhibits only moderately elevated *Δdfi* values (Fig. 3). T69A is the only mutation also flagged as red, whereas the 10 remaining sites for which *Δdfi* values could be calculated appear to transmit the perturbation elsewhere. Two substitutions, R194C and R216H, are found adjacent to red segments in the primary sequence, whereas several others are located more distant from red and blue sites and may play a compensatory role. For example, although the deletion of Tyr217 disrupts a van der Waals contact with Asp74, the D74H substitution compensates by providing H-bonding to LEA Ser217. Similarly, T104R provides electrostatic stabilization to same-chain Glu90. However, T104R also interacts with blue Gly20 across the antiparallel interface, generating the only van der Waals contact of a mutational site with a blue residue. In addition, T104R interacts with neighboring-chain Thr17 and Gly122, thereby enlarging the contact surface. Last, the M154T substitution replaces a solvent-exposed hydrophobic side chain with a hydrophilic one, providing fold stabilization.

Discussion

Dynamics-driven evolution of new function does not require chain rearrangements. In this work, we have uncovered the evolutionary pathway of irreversible green-to-red photoconversion of suborder *Faviina* of the GFP protein family tree. We have solved five high-resolution crystal structures of laboratory-constructed variants that recapitulate phenotypic change along this phylogenetic branch. To aid in the interpretation of static structures, we have subjected the common green ancestor (ALL-Q62H)

and the least evolved ancestor (LEA) to a computational analysis involving MD and PRS methods. A residue-by-residue comparison of the dynamic flexibility index (*dfi*) has allowed us to map sites with pronounced mobility changes onto the protein fold. In our computational approach, both main chain and side chain dynamics are explicitly taken into account. The emerging model suggests that regions of increased softness and regions of increased firmness cluster in opposite corners of the β-barrel (Fig. 5a). Apparently, collective protein motions traverse the cylindrical fold, providing the dynamical framework for a photochemical activation step that yields a red-emitting chromophore. These results imply that a stiff active site is not necessarily a prerequisite for catalysis, in spite of the fact that knob-like regions frequently co-localize with functional groups in the active site (Gerek et al., 2013). The substantial reduction in thermal stability observed for the evolved, photoconversion-competent variants underscores the significance of global dynamics as the decisive parameter in phenotypic change.

This work provides a particularly striking example of dynamics-driven protein evolution, as the process does not appear to be coupled to backbone rearrangements or interactions with allosteric effectors. The GFP single-domain β-barrel fold exhibits a high degree of fold symmetry enforced by extensive tertiary contacts, a feature that sets this system apart from the dynamics-driven evolution of α-helical or α/β proteins with mobile domains or subdomains that may undergo hinge bending motions, loop rearrangements or helical shifts (Ortlund et al., 2007; Schrank et al., 2009; Glembo et al., 2012; Bhabha et al., 2013; Risso et al., 2013). Strikingly, the relocation of a rigidified "knob" region across the cylindrical GFP fold installs the framework necessary for new function, in spite of the severe restriction placed on backbone conformation by a highly connected, closed topology. The corollary of this notion is that in the absence of functional divergence, critical low-frequency normal modes should be conserved in spite of sequence variations, in line with computational results reported some years ago for several motor protein families (Zheng et al., 2006).

We propose that hinge migration provides a more general mechanism for phenotypic change in dynamics-based evolutionary processes. The emerging model embodies a tightly coupled relationship between anchoring and breathing regions that can be tuned by mutations elsewhere in the protein scaffold. Significantly, the rigidified corner that traces about half of the cylinder rim does not harbor any of the 13 mutational sites. In combination, these loop regions decorate the exposed ridge in proximity to the subunit contact surfaces, and serve as a built-in dampening module at the "top right" corner of the barrel (Fig. 5c). Fold stabilization is provided by allosteric interactions with regions of increased deformability at the far face of the chromophore diagonally across the barrel. This feature appears to be critical, as the interior T69A substitution near the "bottom left" corner of the barrel (Fig. 5c) causes substantial thermal destabilization of the protein (Fig. 1). Not surprisingly, this substitution has been shown to be indispensable for efficient photoconversion, and its effect must be compensated for by rigidification elsewhere.

The scattered distribution of mutational sites (Fig. 5cd) warrants that transmission must occur by means of several allosteric pathways, or maybe by a complex allosteric network that involves a large part of the protein fold (del Sol et al., 2009). Indeed, preliminary pathway calculations indicate that there is no single route of communication between mutational sites and the blue knob region. Instead, we speculate that numerous coexisting transmission paths converge onto the knob from different directions, and suggest that the propagation of local stress involves a network of correlated motions that yields firmness at the top ridge for adequate protein stability.

Fundamentally, the evolutionary mechanism described here is closely related to the purely dynamics-based allosteric regulation reported for the catabolite activator and LovD proteins (Popovych et al., 2006; Jimenez-Oses et al., 2014). However, in the present work, we establish a direct experimental link between natural selection and collective dynamics in the absence of substrate, product or effector

binding. Although in our system, the native assembly remains intact, the tetramer does not appear to play a critical role in green-to-red photoconversion (Habuchi et al., 2008).

Functional, epistatic and compensatory mutations promote catalytically important active site flexibility. In the evolved variant, different types of substitutions appear to work synergistically to provide access to all catalytic steps. Functional mutations include Q62H to supply the imidazole building block for the red chromophore, T69A to allow for chromophore twisting (Fig. 2a), and Y116N to facilitate Q62H side chain rotations (Fig. 1c). Epistatic mutations include V157I to direct chromophore distortions towards His193, S105N to anchor the chromophore's imidazolinone ring, and A60V to optimize packing of the helical turn preceding the chromophore. The remaining set of epistatic mutations R194C, R216H, ΔY217, and M218G dislodge the C-terminal clasp, thereby increasing flexibility near the chromophore's phenolic group (Fig. 2bc). Last, compensatory mutations occur at sites remote from catalysis and provide thermal stabilization. M154T improves side chain solvation, D74H provides H-bonding to Ser217 at the base of the C-tail, and T104R forms an intramolecular salt bridge. T104R is the only substitution in van der Waals contact with a blue residue, while also fortifying the antiparallel interface and communicating with the chromophore-bearing pocket.

Based on the pdb database, C-terminal disorder appears to be a conserved feature in Kaede-type proteins, providing a connection between C-peptide dynamics and photochemistry. In all LEA variants, a narrow solvent channel leads to the chromophore (Fig. S3), as reported for several other GFP-like proteins, such as mTFP0.7 (Henderson et al., 2007) and Dronpa (Andresen et al., 2007; Stiel et al., 2007). These proteins exhibit photoswitching between bright and dark states by means of chromophore isomerization, with the dark-state chromophore adopting a non-coplanar, less ordered *trans* conformation with concomitant loss of His193 π-stacking and repositioning of Arg66 and Val157. Significantly, these structural changes cannot proceed without transient disruption of the ionic bond

between His193 and Glu211. Although Ile157 has been shown to hinder chromophore distortions in some systems (Adam et al., 2011), in LEA, V157I may serve to prevent unproductive chromophore isomerization reactions. NMR experiments on Dronpa have demonstrated that part of the β-barrel exhibits increased dynamic motions along segments of strands #7 and #10 lining the solvent channel (Mizuno et al., 2008). As Kaede-type proteins and Dronpa have evolved along the same clade, photoswitching and photoconversion may share common physical characteristics, such as chromophore conformational dynamics and solvent access. These features may be contrasted with the evolution of DsRed along a different branch of the protein family tree, which appears to require a highly structured, firm chromophore-bearing pocket (Wachter et al., 2010).

Reverse protonation of His293 and Glu211 may enhance light-induced chromophore twisting to activate catalytic groups for concerted β-elimination. A more malleable protein scaffold around the chromophore has apparently led to the successful exploration of new chemistry to generate a red-emitting fluorophore. Here, we propose that chromophore twisting in the electronically excited state is facilitated by the increased mobility of Ser142 and His193, residues in direct contact with the phenolic end of the chromophore (Figure 6). Although a substantial energetic barrier exists for displacement of these residues due to the His193-Glu211 salt bridge, the reverse-protonated state may allow such disruption (steps 1 and 2) and therefore constitute the catalytically competent state (Karplus and Pearson, 1997). Recently, we have assigned apparent pK_a values of 7.5 and 4.5 to His193 and Glu211 in LEA (Kim et al., 2014), indicating that at pH 6, only 0.1 % ($10^{-\Delta pKa}$) of the protein population harbors neutral side chains (Fig. 6a, step 1), in line with a photoconversion quantum yield of 10^{-3} to 10^{-4} (Ando et al., 2002; Tsutsui et al., 2005; Kim et al., 2014). We propose that subsequent rotation of the His62 C_α-C_β bond positions the imidazole $N_{\delta1}$ appropriately to accept a proton from Glu211 (step 3). The ensuing imidazolium cation may activate His62-C_β for proton abstraction by the Glu211 carboxylate transiently

lacking electrostatic stabilization (step 4). According to this scenario, the most plausible mechanism for bond scission appears to be of concerted nature as proposed some years ago (Nienhaus et al., 2005), as the simultaneous transfer of a proton from His62 $N_{\varepsilon 2}$ to the leaving group circumvents the formation of a highly unstable carbanion (step 4). In support of this notion, X-ray structures reported for the Kaede-type protein KikGRX include light-induced His62 rotamer states that are sterically well-poised for Glu211-to-His62 proton transfer (Tsutsui et al., 2009). We propose that this mechanism provides, for the first time, a plausible path to a light-activated, transient intermediate harboring both a highly basic Glu211 carboxylate and an acidic His62 β-carbon, consistent with the expected His62 pK_a depression in an apolar micro-environment (Wachter et al., 2010). Subsequent relaxation to equilibrium provides red fluorescence (step 5). Accordingly, it may be possible to increase the photoconversion quantum yield by genetic engineering to narrow the pK_a gap between His193 and Glu211, thereby increasing the catalytically competent population.

Conclusions. Here, we provide a direct connection between natural protein evolution and migration of a rigidified knob to a site distant from the active site. This relationship could not be extracted by a static structural comparison of ancestral and evolved variants. The light-activation of the catalytic process appears to be based solely on chain flexibility, rather than optimized positioning of catalytic groups. Apparently, a distributed set of mutations causes a series of perturbations that propagate through the protein matrix, ultimately converging on a distant location for compensatory stiffening. We propose that this scenario may provide a more general platform for the evolutionary tuning of active site motions.

Experimental Procedures

Protein expression, purification, crystallization and X-ray data collection. C-terminally 6His-tagged proteins were expressed in *E. coli* BL21(DE3) using the plasmid pGEM-T (Promega), and purified by Ni-NTA chromatography. Exposure to room light was avoided, and preparations were stored in the dark at -80°C. Crystals were grown by hanging drop vapor diffusion at 4 ^{o}C (ALL-GFP) or 20 ^{o}C (all others) in the dark, with reservoir solutions of Tris-HCl (pH 7.5-7.7), $MgSO_4$, PEG4000, 15% glycerol (see Supplementary Methods). Crystals were flash-frozen in liquid N_2, and all X-ray diffraction data were collected at 100°K, either in-house using an R-AXIS IV++ image plate detector, Rigaku RU200HB generator and osmic confocal mirrors, or at the Advanced Photon Source (Argonne National Laboratory) beamline ID-19 using an ADSC-Q315R detector, or at the Advanced Light Source (Lawrence Berkeley National Laboratory) beamline 4.2.2 using a NOIR-1 detector.

Structure determination and refinement. Indexing and integration were carried out with MOSFLM (Leslie, 1999), and scaling and merging with SCALA in CCP4 (1994). Structures were solved by molecular replacement using PHASER (McCoy et al., 2007), employing the A-chain of KikG as a search model (PDB ID: 1XSS) (Tsutsui et al., 2005). Auto-tracing was carried out with ARP/wARP (Langer et al., 2008), and side chain positions were adjusted in COOT (Emsley and Cowtan, 2004). R_{free} calculations were carried out with 5% of the data. Rigid body, positional and B-factor refinement was carried out using REFMAC with NCS restraints during early phases of refinement if warranted (Murshudov et al., 1997). The chromophore was modeled by using coordinates from cmFP512 (PDB ID: 2C9J) (Nienhaus et al., 2006), and refined against an appropriate target geometry (Elsliger et al., 1999).

Optical properties, differential scanning fluorimetry and dynamic light scattering (DLS). Absorbance, fluorescence, chromophore pK_a and fluorescence quantum yield determinations were carried out as described, using green protein pools kept in the dark (Kim et al., 2014)**.** For thermal

stability measurements, fluorescence was monitored in a 96-well sealed plate using a LightCycler 480 (Roche Diagnostics, Indianapolis, IN) with λ_{ex} = 483 nm and λ_{em} = 533 nm (40^{o}C to 99^{o}C). The apparent mid-point for protein denaturation ($T_{m\text{-}app}$) was determined manually by recording the temperature at which the fluorescence intensity was half-maximal. DLS experiments were performed using a DynaPro NanoStar instrument (Wyatt Technology Corp., Santa Barbara, CA) at 4 °C and 25 °C. Scattering profiles were obtained at protein concentrations of 0.25, 0.5, 1.5, and 5 mg/ml in 50 mM HEPES pH 7.9, 1 mM EDTA, 20 mM or 150 mM NaCl. The hydrodynamic radius was estimated assuming a spherical particle shape.

Computational methods: *Molecular Dynamics***.** Molecular dynamics (MD) simulations were performed on the All-Q62H and LEA tetrameric structures using the NAMD package (Phillips et al., 2005) with the AMBER FF99SB force field (Hornak et al., 2006) and a Generalized Born implicit solvent model (Onufriev et al., 2004). Geometry optimization of the isolated chromophore in its anionic ground state was performed through a quantum chemical calculation employing a restricted Hartree–Fock method and the 6-31G(d) basis set using GAMESS (Schmidt et al., 1993). *Perturbation Response Scanning (PRS) and Dynamic Flexiblity Index (dfi) calculations.* PRS relies on sequentially applying random external forces to individual residues to induce a perturbation that cascades through the residue interaction network. A Brownian kick is applied to residue *j*, and the ensuing displacement of all remaining residues *i* is computed. Upon iteration, the mean square fluctuation is calculated for each residue *i* in response to perturbation at *j* (Ikeguchi et al., 2005). The differential flexibility index (*dfi*) is determined by normalization of the mean square fluctuation of residue *i* with respect to the average mean square fluctuation of all sites. To compare GFP-like proteins with similar backbone structures but differing residue identities, the canonical coarse-grained Elastic Network Model (ENM) was replaced with all-atom molecular dynamics (MD) simulations, such that the covariance matrix derived from the

MD trajectory could be utilized in PRS calculations rather than the inverse of the Hessian matrix. For details, see Supplemental Information.

Supplemental Information

Supplemental Information includes detailed methods, 3 tables and 5 figures.

Author Contributions

H. K. performed biochemical, optical and crystallographic experiments; T. Z. performed computational experiments; C. M. performed thermostability experiments; K. D. performed DLS experiments; T. J. G. performed expression and crystallization experiments; L. C. performed crystallization screens; R. F. collected synchrotron data; M. V. M. provided plasmids and supervised thermostability experiments; S. B. O. supervised computational experiments; R. M. W. developed research strategy, designed experiments, analyzed data and wrote the manuscript.

Accession numbers

Atomic coordinates and structure factors have been deposited into the RCSB Protein Data Bank. ALL-GFP-old, 4DXI; ALL-Q62H, 4DXM; LEA-X6, 4DXO; LEA-X121, 4DXP.

Acknowledgements

This work was supported by NSF Award No. MCB-0615938 to R. M. W., and by NIH Award No. U54 GM094599 to R. M. W. and S. B. O. Crystallographic data were collected at the Advanced Light Source (ALS) beamlines 4.2.2, 8.2.1 and 8.2.2, and at the Advanced Photon Source (APS) beamline 19 ID. ALS is supported by the US Department of Energy (DOE) under Contract No. DE-AC02-

05CH11231. The Argonne National Laboratory at APS is operated by the University of Chicago Argonne, LLC, for the US DOE Office of Biological and Environmental Research, under contract DE-AC02-06CH11357.

REFERENCES

Adam, V., Moeyaert, B., David, C.C., Mizuno, H., Lelimousin, M., Dedecker, P., Ando, R., Miyawaki, A., Michiels, J., Engelborghs, Y., and Hofkens, J. (2011). Rational design of photoconvertible and biphotochromic fluorescent proteins for advanced microscopy applications. Chemistry & Biology *18*, 1241-1251.

Adam, V., Nienhaus, K., Bourgeois, D., and Nienhaus, G.U. (2009). Structural basis of enhanced photoconversion yield in green fluorescent protein-like protein Dendra2. Biochemistry *48*, 4905-4915.

Alieva, N.O., Konzen, K.A., Field, S.F., Meleshkevitch, E.A., Hunt, M.E., Beltran-Ramirez, V., Miller, D.J., Wiedenmann, J., Salih, A., Matz, M.V. (2008). Diversity and evolution of coral fluorescent proteins. PLoS ONE 3, e2680.

Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H., and Miyawaki, A. (2002). An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. Proc. Natl. Acad. Sci. USA *99*, 12651-12656.

Andresen, M., Stiel, A.C., Trowitzsch, S., Weber, G., Eggeling, C., Wahl, M.C., Hell, S.W., and Jakobs, S. (2007). Structural basis for reversible photoswitching in Dronpa. Proc. Natl. Acad. Sci. U.S.A. *104*, 13005-13009.

Atilgan, C., and Atilgan, A.R. (2009). Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. PLoS Comput. Biol. *5*, e1000544.

Atilgan, C., Gerek, Z.N., Ozkan, S.B., and Atilgan, A.R. (2010). Manipulation of conformational change in proteins by single-residue perturbations. Biophys. J. *99*, 933-943.

Bahar, I., Lezon, T.R., Yang, L.W., and Eyal, E. (2010). Global dynamics of proteins: bridging between structure and function. Annu. Rev. Biophys. *39*, 23-42.

Bhabha, G., Ekiert, D.C., Jennewein, M., Zmasek, C.M., Tuttle, L.M., Kroon, G., Dyson, H.J., Godzik, A., Wilson, I.A., and Wright, P.E. (2013). Divergent evolution of protein conformational dynamics in dihydrofolate reductase. Nat. Struct. Mol. Biol. *20*, 1243-U1262.

Collaborative Computational Project No. 4 (1994). The CCP4 Suite: Programs for protein crystallography. Acta Crystallogr. *D50*, 760-763.

del Sol, A., Tsai, C.J., Ma, B.Y., and Nussinov, R. (2009). The origin of allosteric functional modulation: Multiple pre-existing pathways. Structure *17*, 1042-1050.

Elsliger, M.-A., Wachter, R.M., Hanson, G.T., Kallio, K., and Remington, S.J. (1999). Structural and spectral response of green fluorescent protein variants to changes in pH. Biochemistry *38*, 5296-5301.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D *60*, 2126-2132.

Field, S.F., Bulina, M.Y., Kelmanson, I.V., Bielawski, J.P., and Matz, M.V. (2006). Adaptive evolution of multicolored fluorescent proteins in reef-building corals. J. Mol. Evol. *62*, 332-339.

Field, S.F., and Matz, M.V. (2010). Retracing evolution of red fluorescence in GFP-like proteins from Faviina corals. Mol. Biol. Evol. *27*, 225-233.

Gerek, Z.N., Kumar, S., Ozkan, S.B. (2013). Structural dynamics flexibility informs function and evolution at a proteome scale. Evol. Appl. *6*, 423-433.

Gerek, Z.N., and Ozkan, S.B. (2011). Change in allosteric network affects binding affinities of PDZ Domains: Analysis through Perturbation Response Scanning. PLoS Comput. Biol. *7*, e1002154.

Glembo, T.J., Thorpe, M.F., Farrell, D.W., Gerek, Z.N., and Ozkan, S.B. (2012). Collective dynamics differentiates functional divergence in protein evolution. PLoS Comput. Biol. *8*, e1002428.

Habuchi, S., Tsutsui, H., Kochaniak, A.B., Miyawaki, A., and van Oijen, A.M. (2008). mKikGR, a monomeric photoswitchable fluorescent protein. PLoS ONE *3*, e3944.

Henderson, J.N., Ai, H.-W., Campbell, R.E., and Remington, S.J. (2007). Structural basis for reversible photobleaching of a green fluorescent protein homologue. Proc. Natl. Acad. Sci. U.S.A. *104*, 6672-6677.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins: Struct., Funct., Bioinf. *65*, 712-725.

Ikeguchi, M., Ueno, J., Sato, M., and Kidera, A. (2005). Protein structural change upon ligand binding: Linear response theory. Phys. Rev. Lett. *94*, 078102.

Jimenez-Oses, G., Osuna, S., Gao, X., Sawaya, M.R., Gilson, L., Collier, S.J., Huisman, G.W., Yeates, T.O., Tang, Y., and Houk, K.N. (2014). The role of distant mutations and allosteric regulation on LovD active site dynamics. Nat. Chem. Biol. *10*, 431-436.

Karplus, P.A., and Pearson, M.A. (1997). 70 Years of crystalline urease: What have we learned? Acc. Chem. Res. *30*, 330-337.

Kennis, J.T.M., van Stokkum, I.H.M., Peterson, D.S., Pandit, A., Wachter, R.M. (2013). Ultrafast Proton Shuttling in Psammocora Cyan Fluorescent Protein. J. Phys. Chem. B *117*, 11134-11143.

Kim, H., Grunkemeyer, T.J., Modi, C., Chen, L., Fromme, R., Matz, M.V., Wachter, R.M. (2013). Acid-base catalysis and crystal structures of a least-evolved ancestral GFP-like protein undergoing green-to-red photoconversion. Biochemistry *52*, 8048-8059.

Langer, G., Cohen, S.X., Lamzin, V.S., Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. *3*, 1171-1179.

Leslie, A.G. (1999). Integration of macromolecular diffraction data. Acta Crystallogr. D *55*, 1696-1702.

McCoy, A., Grosse-Kunstleve R, Adams P, Winn M, Storoni L, Read R (2007). Phaser crystallographic software. J. Appl. Cryst. *40*, 658-674.

Mizuno, H., Mal, T.K., Walchli, M., Kikuchi, A., Fukano, T., Ando, R., Jeyakanthan, J., Taka, J., Shiro, Y., Ikura, M., Miyawaki, A. (2008). Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. Proc. Natl. Acad. Sci. USA *105*, 9227-9232.

Moeyaert, B., Bich, N.N., De Zitter, E., Rocha, S., Clays, K., Mizuno, H., Van Meervelt, L., Hofkens, J., Dedecker, P. (2014). Green-to-red photoconvertible Dronpa mutant for multimodal super-resolution fluorescence microscopy. Acs Nano *8*, 1664-1673.

Murshudov, G.N., Vagin, A.A., Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelyhood method. Acta Crystallogr. D*53*, 240-255.

Nienhaus, K., Nienhaus, G.U., Wiedenmann, J., Nar, H. (2005). Structural basis for photo-induced protein cleavage and green-to-red conversion of fluorescent protein EosFP. Proc. Natl. Acad. Sci. USA *102*, 9156-9159.

Nienhaus, K., Renzi, F., Vallone, B., Wiedenmann, J., Nienhaus, G.U. (2006). Exploring chromophore-protein interactions in fluorescent protein cmFP512 from Cerianthus membranaceus: X-ray structure analysis and optical spectroscopy. Biochemistry *45*, 12942-12953.

Onufriev, A., Bashford, D., Case, D.A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins: Struct., Funct., Bioinf. *55*, 383-394.

Ortlund, E.A., Bridgham, J.T., Redinbo, M.R., Thornton, J.W. (2007). Crystal structure of an ancient protein: Evolution by conformational epistasis. Science *317*, 1544-1548.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781-1802.

Popovych, N., Sun, S.J., Ebright, R.H., Kalodimos, C.G. (2006). Dynamically driven protein allostery. Nat. Struct. Mol. Biol. *13*, 831-838.

Pouwels, L.J., Zhang, L., Chan, N., Dorrestein, P., Wachter, R.M. (2008). Kinetic isotope effect studies on the de novo rate of chromophore formation in fast- and slow-maturing GFP variants. Biochemistry *47*, 10111-10122.

Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A., Sanchez-Ruiz, J.M. (2013). Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. J. Am. Chem. Soc. *135*, 2899-2902.

Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.H., Koseki, S., Matsunaga, N., Nguyen, K.A., Su, S.J., Windus, T.L., Dupuis, M., Montgomery, J.A. (1993). General atomic and molecular electronic-structure system. J. Comput. Chem. *14*, 1347-1363.

Schrank, T.P., Bolen, D.W., Hilser, V.J. (2009). Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. Proc. Natl. Acad. Sci. USA *106*, 16984-16989.

Stiel, A.C., Trowitzsch, S., Weber, G., Andresen, M., Eggeling, C., Hell, S.W., Jakobs, S., Wahl, M.C. (2007). 1.8 A bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants. Biochem. J. *402*, 35-42.

Tokuriki, N., Tawfik, D.S. (2009). Protein dynamism and evolvability. Science *324*, 203-207.

Tsutsui, H., Karasawa, S., Shimizu, H., Nukina, N., Miyawaki, A. (2005). Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. Embo Rep. *6*, 233-238.

Tsutsui, H., Shimizu, H., Mizuno, H., Nukina, N., Furuta, T., Miyawaki, A. (2009). The E1 mechanism in photo-induced beta-elimination reactions for green-to-red conversion of fluorescent proteins. Chemistry & Biology *16*, 1140-1147.

Ugalde, J.A., Chang, B.S.W., Matz, M.V. (2004). Evolution of coral pigments recreated. Science 305: 1433-1433

Wachter, R.M., Watkins, J.L., Kim, H. (2010) Mechanistic diversity of red fluorescence acquisition by GFP-like proteins. Biochemistry *49*, 7417-7427.

Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Roecker, C., Salih, A., Spindler, K.-D., Nienhaus, G.U. (2004). EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. Proc. Natl. Acad. Sci. USA *101*, 15905-15910.

Yarbrough, D., Wachter, R.M., Kallio, K., Matz, M.V., Remington, S.J. (2001). Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-A resolution. Proc. Natl. Acad. Sci. USA *98*, 462-467.

Zhang, L., Patel, H.N., Lappe, J.W., Wachter, R.M. (2006) Reaction progress of chromophore biogenesis in green fluorescent protein. J. Am. Chem. Soc. *128*, 4766-4772.

Zheng, W., Brooks, B.R., Thirumalai, D. (2006). Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. Proc Natl Acad Sci U S A *103*, 7664-7669.

Zheng, W.J., Brooks, B.R., Doniach, S., Thirumalai, D. (2005). Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. Structure *13*, 565-577.

FIGURE LEGENDS

Figure 1. **(a)** Thermostability of reconstructed ancestral proteins. $T_{m\text{-}app}$ values were determined by irreversible loss of intrinsic fluorescence. Plotted is the average of duplicate determinations vs. protein concentration. **(b)** $2F_o\text{-}F_c$ electron density map of the ALL-Q62H chromophore-bearing pocket contoured at 1.5 σ. **(c)** Overlay of the X-ray structures of ALL-Q62H (grey) and LEA (green). Yellow dashed lines: LEA H-bonds. Red dashed line: close contact. See also Figures S1 and S2.

Figure 2. **(a)** Superimposition of the X-ray structures of ALL-Q62H (grey) and LEA (red and cyan), depicting the T69A substitution and Arg66 side chain reorientation. Residues with increased flexibility (elevated *Δdfi*) in LEA are colored red. **(b, c)** X-ray structural models of the C-termini of **(b)** ALL-Q62H and **(c)** LEA. Adjacent subunits are colored cyan and yellow (perpendicular interface). Dashed lines represent hydrogen or ionic bonds. See also Figure S3.

Figure 3. **(a)** MD calculations of the average chromophore RMSF in ALL-Q62H and LEA, calculated for the last 100 ns with a sliding window of 5 ns. **(b)** The dynamics profile (*%dfi*) of ALL-Q62H and LEA (top and middle row), and the difference between the two (*Δdfi*) (bottom). Red: Residues with increased flexibility in LEA. Blue: residues with decreased flexibility in LEA. Green: chromophore-forming residues 62, 63 and 64. Vertical dashed lines: mutational sites. Residues colored red: 67-71, 142, 187-193 and 213-215. Residues colored blue: 19-22, 50, 97-98, 125, 127-129, 165 and 167. See also Figure S4 and S5.

Figure 4. Comparison of the dynamics profile of ALL-Q62H and LEA. Residues are rainbow-colored according to their *%dfi* values, with rigid regions in blue tones and flexible regions in red tones.

Residues with significantly modified flexibility are shown as space-filling models (main chain and side chain atoms). **(a)** Side view of ALL-Q62H; **(b)** Side view of LEA.

Figure 5. Top-scoring 15% of residues with large Δ*dfi* values mapped onto the structure of ALL-Q62H. Red, elevated flexibility and blue, decreased flexibility in LEA. Green, chromophore. **(a, b)** Main-chain atoms of red and blue residues are shown as spheres. **(c,d)** Main-chain atoms of the mutational sites are shown as grey spheres.

Figure 6. Proposed photoconversion mechanism. **(a)** Reverse protonation promotes light-activation of Glu211 and His62 by allowing chromophore distortions (steps 1 and 2). His62 rotamer adjustements facilitate proton shuttling (steps 2 and 3), and C_β-deprotonation occurs concerted with leaving group ejection (step 4). **(b)** Cartoon depicting β-barrel deformations proposed to occur during photoconversion. Charge states of Glu211 and His62: blue, positive; red, negative; grey, neutral.

Tables

Table 1. Crystallographic data collection statistics. See also Tables S1, S2 and S3.

	ALL-GFP	ALL-Q65H	LEA-X6	LEA-X121
space group	C2	C2	I222	I222
chains per AU	4	4	1	1
a (Å) b (Å) c (Å) β (°)	106.49 93.06 88.99 92.72	106.28 92.71 89.06 93.18	45.65 78.34 119.43 90	49.04 79.18 118.04 90
Detector	R-AXIS IV++	NOIR-1	R-AXIS IV++	ADSC Q315
Wavelength (Å)	1.5418	0.9998	1.5418	0.9792
Resolution (Å)	28.35 – 1.60	29.67 – 1.40	39.16 – 2.50	34.05 – 1.75
High-res. shell (Å)	1.69 – 1.60	1.48 – 1.40	2.64 – 2.50	1.84 – 1.75
Total observations	264,463	598,321	39,140	84,219
Unique reflections	111,934	169,035	7,913	23,358
Redundancy[a]	2.4 (2.1)	3.5 (3.5)	4.8 (4.9)	3.6 (3.7)
Average I/σ[a]	13.4 (1.8)	6.1 (1.8)	16.2 (4.1)	7.6 (2.3)
Completeness (%)		99.9(100.0)	98.8 (98.4)	99.3 (99.9)
$R_{merge\ (\%)}$[b]	4.7(42.8)	7.6 (37.8)	5.8 (36.3)	8.2 (38.1)

[a]Values within parentheses refer to the high-resolution shell.

[b]$R_{merge} = \Sigma|I_{hkl} - \langle I \rangle| / \Sigma\langle I \rangle$, where $\langle I \rangle$ is the average of individual measurements of I_{hkl}.

Table 2. Crystallographic refinement statistics.

	ALL-GFP	ALL-Q65H	LEA-X6	LEA-X121
Resolution range (Å)	28.12 – 1.60	29.67 – 1.40	30.55 – 2.50	34.05 – 1.75
Reflections	106,316	160,594	7,527	22,027
R_{cryst} (%)[c]	17.3	14.3	19.7	17.1
R_{free} (%)[c]	21.0	17.9	24.7	22.6
rmsd bonds (Å)	0.011	0.011	0.018	0.016
rmsd angles (°)	1.402	1.492	1.758	1.629
avg B-factors (Å^2)	11.4	10.4	31.5	20.8
Total atoms	8,082	8,576	1,775	1,960
Residues modeled: chain A chain B chain C chain D	 2-221 2-225 2-223 2-221	 2-224 2-224 2-231 2-231	 4-218	 2-220

[c] R_{cryst} and $R_{free} = \Sigma_h (||F(h)_{obs}| - |F(h)_{calc}||) / |F(h)_{obs}|$ for reflections in the working and test sets (5% of all data).

Figure 1

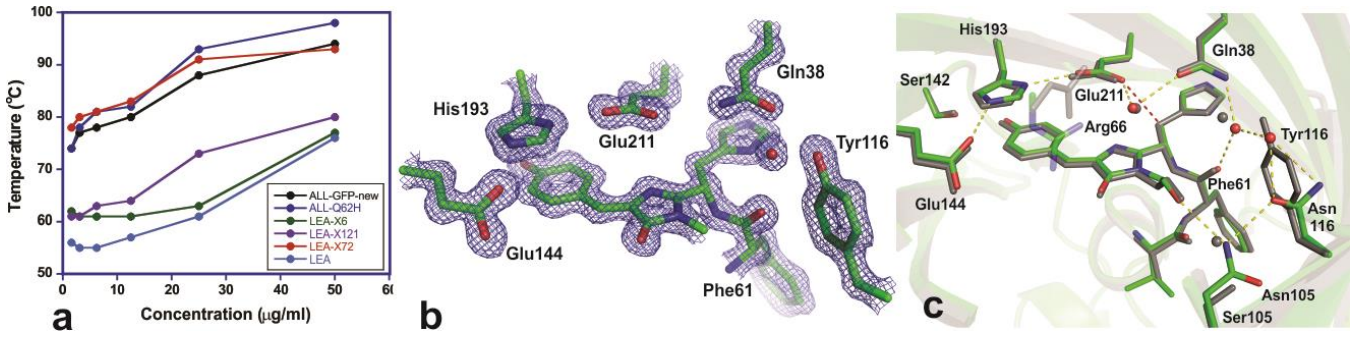


Figure 2

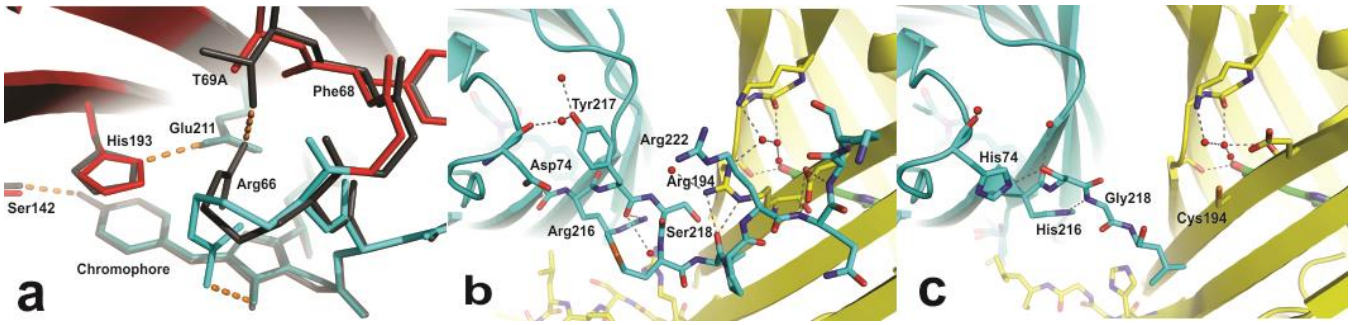


Figure 3

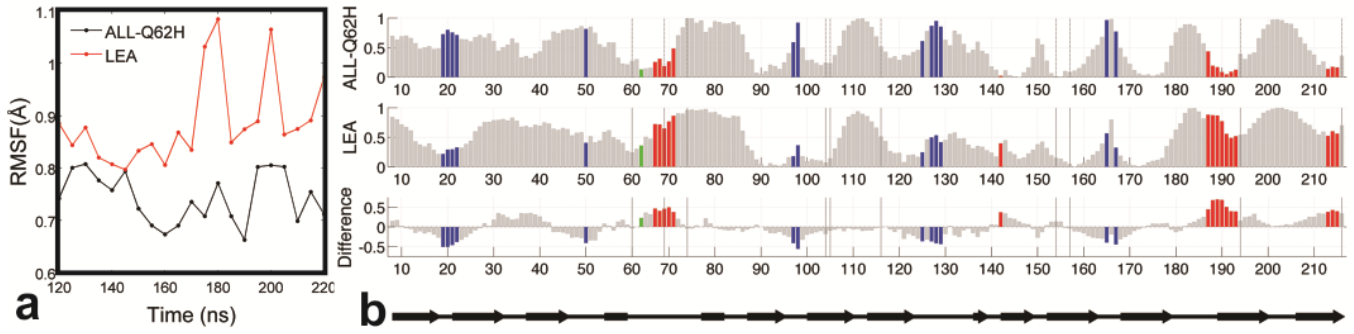


Figure 4

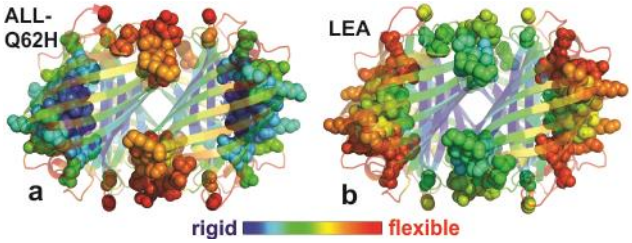


Figure 5

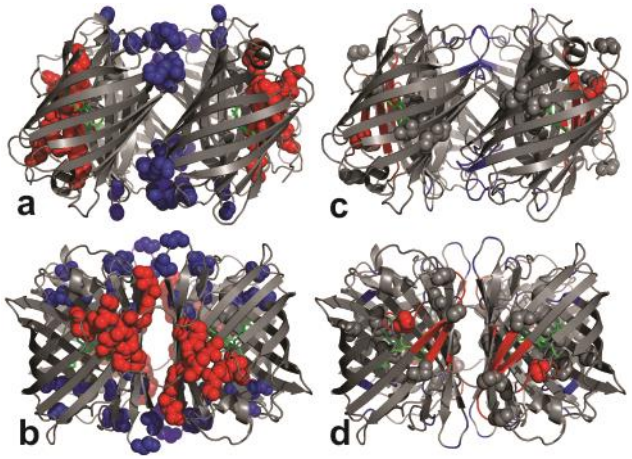
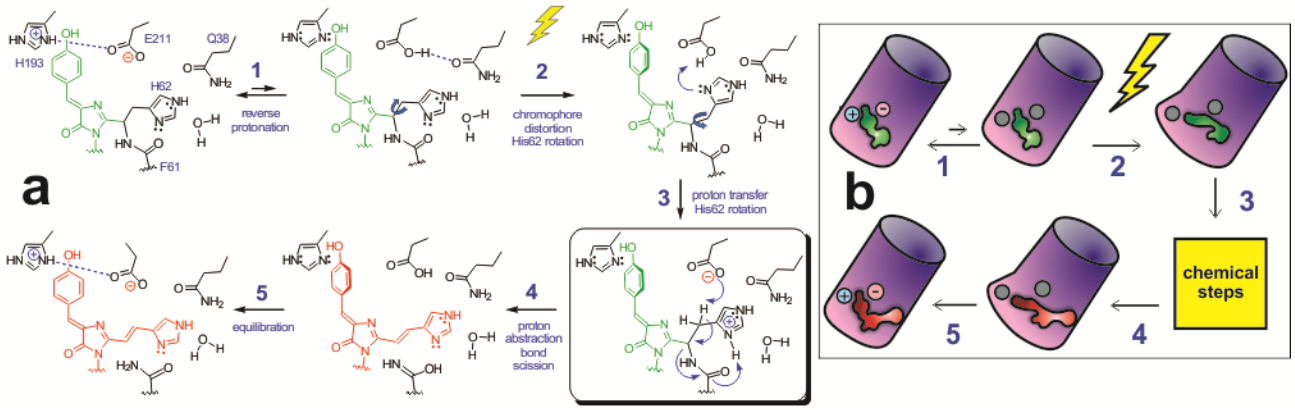


Figure 6


a
E211
Q38
H193
H62
F61
1
reverse protonation
2
chromophore distortion His62 rotation
3
proton transfer His62 rotation
4
proton abstraction bond scission
5
equilibration
b
1
2
3
4
5
chemical steps