Published in final edited form as:

Chem Phys. 2014 January 31; 429: 5–11. doi:10.1016/j.chemphys.2013.11.015.

Structural Consequences of Chromophore Formation and Exploration of Conserved Lid Residues amongst Naturally Occurring Fluorescent Proteins

Matthew H. Zimmer^{a,b}, Binsen Li^a, Ramza S. Shahid^a, Paola Peshkepija^a, and Marc Zimmer^{a,*}

^aChemistry Department, Connecticut College, New London, CT06320, USA

^bDepartment of Chemistry, University of Oxford, Oxford OX1 3QZ, United Kingdom

Abstract

Computational methods were used to generate the lowest energy conformations of the immature precyclized forms of the 28 naturally occurring GFP-like proteins deposited in the pdb. In all 28 GFP-like proteins, the beta-barrel contracts upon chromophore formation and becomes more rigid. Our prior analysis of over 260 distinct naturally occurring GFP-like proteins revealed that most of the conserved residues are located in the top and bottom of the barrel in the turns between the β -sheets.(1) Structural analyses, molecular dynamics simulations and the Anisotropic Network Model were used to explore the role of these conserved lid residues as possible folding nuclei. Our results are internally consistent and show that the conserved residues in the top and bottom lids undergo relatively less translational movement than other lid residues, and a number of these residues may play an important role as hinges or folding nuclei in the fluorescent proteins.

Introduction

Crystal jellyfish and their green fluorescent proteins (GFP) have been floating in the ocean for more than 160 million years,(2, 3) before a quartet of curious scientists, fascinated by pinpricks of their green light, began unlocking its potential. Now, GFP is the microscope of the 21st century. In technicolor, it allows us to see things we have never been able to see before, thereby completely changing the way we approach science and medicine. (4–8)

The importance and widespread use of applications based on fluorescent proteins has fuelled a search for new FPs in both nature and mutant space. Over 260 distinct naturally occurring GFP-like proteins are currently known. GFP-like fluorescent proteins (FPs) have been found in marine organisms ranging from chordates (e.g. amphioxus) to cnidarians (e.g. corals and sea pansies).

© 2013 Elsevier B.V. All rights reserved

*Box 5624, Connecticut College, 270 Mohegan Ave., New London, CT06320, USA, mzim@conncoll.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Supporting Information. Tables listing RMSDs after structural overlaps, dihedral variations and RMSDs of the MD simulations, hinge-like behaviour and average square fluctuations from ANM and figures listed in the text are available free of charge via the Internet at <http://pubs.acs.org>.

An analysis of the structures of the GFP-like proteins in the PDB revealed that most of the conserved residues are located in the top and bottom of the barrel in the turns between the β -sheets.(1) Herein we have used computational methods to examine whether the lid residues may be conserved because they have a critical function in the folding of the β -barrel.

According to the funnelled energy landscape theory there are many folding pathways for a protein to adopt, although a small number of them dominate the folding process.(9, 10) In order to form a correctly folded protein these energy funnels have to be robust under a large variety of conditions.

It has long been known that active site residues are commonly conserved, however recent studies have shown that evolutionary conservation and structural dynamics are also strongly linked.(11, 12) Although the protein folding pathways are minimally affected by most mutations,(13) folding nuclei that are critically important in helping proteins adopt their three dimensional conformations are highly conserved.(14) Local perturbations or interference of hinge sites can give rise to allosteric effects or even disrupt the entire cooperativity of the functional motions of a protein, and therefore it is not surprising that these sites are also conserved.(15)

GFP-like proteins fold into a distinctive beta barrel shape composed of 11 β -sheets surrounding a central alpha helix, which contains the chromophore, see Figures 1 and 2. In *Aequorea victoria* GFP (avGFP), the chromophore is formed by an autocatalytic cyclization of the tripeptide S65Y66G67 fragment. Folding of the tertiary structure is fast, though chromophore formation and fluorescence is not observed until 90 minutes to 4 hours after protein synthesis.(16–18) The folding of GFP exhibits hysteresis due to the decreased flexibility of the chromophore vs. its immature analogue(19–21), and the compaction of the β -barrel upon chromophore formation.(22) All signs of hysteresis disappear in mutants that do not form the chromophore.(23) Single-molecule fluorescence(24), mechanical(25) studies, and simulations thereof,(26) NMR(27), denaturing and renaturing experiments(28) and coarse-grained molecular simulations(29) have been used to examine folding in fluorescent proteins. Reddy et al. have combined the results from their coarse grained simulations with reported experimental observations(25, 27, 28, 30) to propose a model for GFP folding.(29) It includes multiple pathways, passing through kinetic and equilibrium intermediates as well as misfolded structures.

In this study, we have used a variety of computational techniques to analyze the structure of GFP-like proteins in both their native and uncyclized, immature states. Our results from using both detailed atomistic simulations and simplified models on proteins structurally similar to the crystal structures bear strong resemblance to the experimental and theoretical unfolding experiments mentioned above, while providing additional information about the conserved residues in the lids of the β -barrel.

Methods

Structure Preparation of Immature Wild-type FPs and Conformational Searches

The coordinates of the crystal structures of all the wild-type GFP-like proteins were obtained from the Protein Data Bank (PDB)(31) - (1GFL(32), 1MOU(33), 1UIS(34), 1XSS(35), 1YZW(36), 1ZGO(37), 1ZUX(38), 2A46(39), 2C9I(40), 2C9J(41), 2DD7(42), 2G3O(36), 2GW3(43), 2IB5(44), 2ICR(45), 2IE2(46), 2OGR(47), 2OJK(45), 2RH7(48), 2WHT(49), 2Z6X(50), 2ZMU(51), 3CGL(52), 3GB3(53), 3H1O(54), 3MGF(55), 3PIB(53), and 3PJ5(53)). The protein preparation workflow(56) and Epik v2.0109(57) were used with hydrogen bond optimization to add hydrogen atoms to protein and solvent atoms as required. The OPLS_2005 force field of MacroModel v9.8107(58) was used.

The starting structure for the immature forms of the fluorescent proteins were calculated by graphically mutating the chromophores of the mature FP crystal structures so that the chromophore forming tripeptide sequences were in the original precyclized form before undertaking a conformational search, Figure 1. Conformational searches were conducted using the combined Monte Carlo torsional variation and low mode method.(59, 60) The flexible dihedral angles of all the side-chains of residues 64, 65, 66, 67 and 68 (1GFL numbering) were randomly rotated by between 0 and 180° and all solvent molecules in an 8.00 Å sphere from residues 64–68 were randomly rotated and translated by between 0 and 1.00 Å in each Monte Carlo (MC) step.(61) Conformational searches were carried out until 500 MC steps were taken without finding new conformations.

Structural Conservation

The MatchMaker extension of Chimera(62) was used to examine the structural conservation of the conserved residues in all wild type fluorescent proteins in the PDB and their computationally determined immature forms. A least-squares fit of pairs of sequenced aligned alpha-carbons was performed with the default settings in Chimera. A Needleman-Wunsch algorithm was used to best match the structure of avGFP with the other 27 wild-type fluorescent proteins. After the superposition of the 28 proteins, a structural sequence alignment was made through Match -> Align subroutine with a 5 Å residue-residue distance cutoff. Finally, Multalign Viewer extension was used to produce sequence alignments together with associated structures.(16, 17)

Molecular Dynamics

The coordinates of the *Aequorea victoria* GFP crystal structure (1gfl)(63) were obtained from the Protein Data Bank (PDB)(31) and prepared as described above. 15,000 MC steps were taken in each search. Structures within 50kJ/mol of the lowest energy minimum were kept, and a usage directed method(60) was used to select structures for subsequent MC steps. Structures found in the conformational search were considered unique if the least squared superimposition of equivalent non-hydrogen atoms found one or more pairs separated by 0.25 Å or more. The lowest energy structure obtained in the search was further subjected to a 5,000 step large scale low mode conformational search.(64, 65)

The final structures obtained from the fully minimized pdb structures and the conformational searches were used to initiate molecular dynamics (MD). MD simulations were carried out in the NPT ensemble at 300K and 1 bar with 1.5 fs steps using Desmond(66). All molecular dynamics calculations used the OPLS_2005 force field and SHAKE constrained hydrogens. 10,418 structures were sampled in each 50 ns MD simulation. Each structure was in an orthorhombic simulation box of 0.15M NaCl and SPC waters,(67) with a 10 Å solvent buffer between the protein surface and the boundary. Minimizations and pre-equilibrium simulations were done using the default Desmond/multisim relaxation protocol.

Fifty nanosecond simulations of avGFP were conducted with the chromophoric Tyr66 in both the neutral and anionic forms. The immature precyclized forms of avGFP were only simulated with the neutral form of the Tyr66, see Figure 1.

Root mean square deviation (RMSD) was calculated on a per residue basis by finding the RMSD of all the backbone atoms in a residue from a particular frame compared to the same atoms in the averaged structure from the trajectory. This RMSD per frame was then averaged over the entire trajectory to produce the final value for a residue.

Separate 50 nanosecond simulations of the $\beta 4$ – $\beta 5$ fragment of wild-type avGFP and its G104A mutant were undertaken using the conditions described above. The fragments consisting of residues 91 to 116 were obtained from the phenolic form of avGFP after adding hydrogens and conducting a conformational search as described above, and then capped using the protein preparation workflow.

Anisotropic Network Model

The normal mode vibrations of 5 wild-type GFPs (PDB codes: 1GFL, 1ZUX, 1MOU, 1UIS, and 2ICR) were analyzed using the Anisotropic Network Model (ANM) web server.⁽⁶⁸⁾ The best correlation between the experimental and theoretical B-factors was obtained by systematically changing the cutoff distance (ranging from 15–21 Å) and distance weights (ranging from 0–3) for interactions between C α atoms, Figure S1.⁽⁶⁹⁾ The cut-off distance and distance weight values were changed and their corresponding Pearson's correlation coefficient values were recorded. Since the wild-type FPs are composed of multiple chains that are structurally similar, only chain A for each of them was used in these studies. B-factor analysis was conducted to compare each of the mean square fluctuations of individual residues of the experimental and theoretical B-factors, while mode analysis was conducted to examine the mean square fluctuations of individual residues in different modes of vibration. To compare the mature and immature proteins' total square fluctuations, we used the ProDy interface with a spring constant of 1.0N/m and a cutoff of 19 Å.⁽⁷⁰⁾

Results and Discussion

Inspection of the twenty most conserved residues in all naturally occurring GFP-like proteins (85 species) has revealed that most of the residues are located in the area comprising the lids of the barrels - residues 40, 75, 89, 91 and 196 on the side of the N and C termini and residues 23, 55, 100–104, 130, 134, 136, and 174 on the other end of the barrel, which is referred to as the top of the barrel, Figure 2. (1) This conservation is despite the fact that the top lid is far removed from the chromophore and has the least overall structure and no alpha helices. We propose that the lid residues may be conserved because they have a critical function in the folding of the β -barrel and that the conserved residues may be acting as hinges. Since it is known that like door hinges, hinge sites in proteins have rotational flexibility, but no translational mobility; and that they are highly conserved,⁽⁷¹⁾ one should be able to distinguish hinge residues in naturally occurring GFP-like proteins by the fact that they are conserved, translationally invariant, yet dihedrally flexible. The structures we have examined, whether mature or immature, are fully folded structures and we therefore expect the hinge residues to be conserved and translationally invariant, but do not expect to see the dihedral freedom that would be more evident in the folding or unfolding of the protein.

We have chosen to examine all naturally occurring GFP-like proteins (with no mutations) that have been crystallized and whose structures are found in the pdb. Since the mature and immature FPs fold differently, but presumably use the same hinge residues, we have used computational methods to generate the uncyclized immature forms of the naturally occurring GFP-like proteins and analyzed them as well. The immature structures were obtained by graphically modifying the chromophore of the 28 wild type fluorescent proteins listed in the methods section to the precyclized, immature chromophore, and running a conformational search to find the lowest energy structure.

Structural Overlap of Naturally Occurring FPs and Their Computationally Generated Immature Analogs

If one assumes that the protein backbone will adopt a low energy conformation in most crystal structures, then the conformational space spanned by the crystal conformations can be equated and be compared (qualitatively not quantitatively) with the potential energy surface of the backbone itself.(72) Therefore a structural alignment was performed using UCSF Chimera to find the RMSD between the α carbons in the 28 naturally occurring GFP-like proteins found in the pdb. Table 1 compares these RMSDs to those obtained for their immature counterparts. The 28 immature and mature structures were matched against the structure of avGFP (pdb code = 1GFL), Figure 3. Graphs of RMSD per residue for both structures can be found in Figure S2.

Table 1 shows that in all subsets of residues examined, the RMSD of the immature structures are higher than those of the mature structures. It is most likely that this is caused by the constriction of the β barrel that occurs as a consequence of chromophore formation. (22) In both the immature and mature structures there is relatively less structural difference between residues located in the beta-sheets than those found in the lid residues(see Figure 3). This is not surprising since the inter-strand hydrogen bonding is responsible for a much more constrained structure between the beta sheets than that found in the more translationally variable turn residues.

The function of most of the conserved residues found outside the lid areas, such as those in the central β barrel and α helix is known. They have been found to play important roles in chromophore formation, (18, 73, 74) which can explain both their conservation and low RMSD. For example, R96, a major catalytic residue, has low RMSDs of 0.44 and 0.60 Å in the mature and immature structures, respectively.

Table 1 reveals that the conserved residues in the top and bottom lids also have relatively lower RMSDs than other lid residues. They are unlikely to play a role in chromophore maturation and we suggest a number of these residues may play an important role as hinges in GFP. These hinges can control the movement of large swaths of proteins through only small local deviations. Turns would be the logical starting place to begin looking for these hinges, and the turn between β 4 and β 5 appears particularly promising, as it consists of residues K101 (1.26 Å immature RMSD), D102 (1.73 Å), D103 (1.93 Å), and G104 (0.69 Å), all of which are conserved.

Molecular Dynamics

Fifty nanosecond molecular dynamics simulations of neutral, anionic and immature avGFP were run to examine the behavior of the conserved lid residues. Unfortunately, due to the computation requirements and the fact that generating the immature FPs was a CPU intensive process, 50 ns simulations were not generated for all 28 wild-type FPs.

The root mean square deviations of the backbone atoms in each residue were calculated as described in the MD section of Methods. Although the molecular dynamics simulation of avGFP and the structural overlap of the 28 naturally occurring FPs are two very dissimilar approaches to examining the role of conserved residues, they produce the same results. The RMSDs for each residue in all three avGFP forms are plotted Figure 4 and listed in Table 2.

Figure 4 gives a graphic overview of many of the trends elaborated upon in the text and tables below. Many of the maxima are positioned between the β sheets, while many of the minima occur at the locations of the conserved residues. According to the data listed in Table 2 the immature structures have consistently higher RMSDs than their corresponding mature structures. We suspect that this is due to the compaction of the β -barrel that occurs

upon chromophore formation – a contraction that is probably one of the driving forces for chromophore formation. The structural analysis of the mature and immature forms of the 28 naturally occurring FPs revealed a similar increase in rigidity between the more flexible immature structures and the more rigid fluorescent forms.

We also observed that in all three structures the RMSDs of the last five β strands ($\beta 7$ – $\beta 11$) were greater than the first six $\beta 1$ – $\beta 6$ (Table 2). These groupings of β sheets were examined as they were also found to move in a concerted fashion in our anisotropic network model analysis described below and because these results may reflect GFP's tendency to unfold from $\beta 7$ – $\beta 11$, which has been observed in other theoretical and experimental studies (see Figure 2).(29)

As observed for the structural overlaps of the FP crystal structures (Table 1), in the avGFP MD simulations the conserved lid residues have lower RMSDs than their non-conserved counterparts, indicating that they may be acting as hinges. Although the simulations were performed on fully folded FPs we also examined the dihedral motion of residues by calculating the standard angular deviation of the Φ and ψ dihedrals on each residue in the immature avGFP structure over the 50 ns trajectory. Table 2 shows that there is more dihedral freedom among the conserved residues in the top of the lid than in any of the other sets sampled. This indicates there may be multiple hinge residues in the top. By simply multiplying the averages of the alignment and MD RMSDs, then dividing by the average of the dihedral deviation, we were able to rank the conserved lid residues in terms of how hinge-like their behaviors are, with lower values indicating potential hinges (Eq. 1). We found that the top three hinge-like residues are 104, 103, and 100, with values of 1.88, 2.36, and 4.82 respectively, see Table S5. All of these residues are in the turn between $\beta 4$ and $\beta 5$. 50 ns MD simulations of two fragments, one wild type and one G104A mutant, composed of these β and the turn between them showed no unfolding tendencies, nor significant differences between them.

$$H = \frac{(M_a + M_R + M_i)(A_m + A_i)}{3(D_\phi + D_\theta)}, \quad \text{Equation 1}$$

where H is a measure of hinge-like behavior, M is the RMSD from the MD simulation, A is the RMSD for structural alignment, D is the dihedral deviation, a is the anionic form, n is neutral and i is immature.

The 21 low energy structures found to be within 50 kJ of the global minimum in a conformational search of the immature 1GFL structure were compared with the lowest energy structure to find the conformational space available to the immature form of avGFP. The results qualitatively match those of the structural overlaps and MD simulations, Figure S3.

Anisotropic Network Model of Computationally Generated Immature Wild Type FPs

The Anisotropic Network Model (ANM) is a powerful normal mode analysis tool that predicts protein dynamics from its native state structure, and thus provides another way to observe the low energy movements of fluorescent proteins. It uses a coarse-grained approach to represent a protein as a network where each node is the C α of a residue. The interactions between these nodes are represented by a single, uniform force constant. This network can then be built into a Hamiltonian matrix, which when decomposed into its eigenvectors provide information on the low frequency vibrations of the protein. These cooperative slow modes and their combinations are the dominant motions of the protein. See Ref. (68) for more detailed descriptions and explicit formulas. ANM has previously been shown by Bahar to have a correlation with mechanical unfolding for GFP.(26)

Calculation of the mean square fluctuation of the first ten modes shows that there is relatively little movement of the β sheets, while the greatest fluctuation comes from the turns between the β sheets. This is the expected behavior, and similar results have been observed by other groups(24) and were also seen in both our structural analyses (Figure 3) and molecular dynamics simulations. Higher modes usually display more global fluctuations involving the β barrel. Since we expect the compaction associated with chromophore formation to be associated with more global fluctuations we measured the total square fluctuation as the sum of the average residue square fluctuations in modes 10–100. As with the various RMSD results described above, for all 28 wild type structures, the immature form has greater movements in each protein as compared to the mature structure (see Table S6). The force constant (γ) was set to 1.0 for each ANM calculation.

More results were obtained by analyzing the inter-residue distance, Figure 5. Elements of secondary structure are clearly visible in the figure. The red diagonal streaks, indicative of low inter-residue fluctuations, across the graph are due to the beta sheets, while the α helix can be seen as the thick horizontal and vertical lines (A in Fig. 5). Turns can also be seen as lighter colored horizontal and vertical lines indicating areas of high inter-residue fluctuation.

Quadrant 1 (Q1) in Figure 5 encompasses the interactions of residues 1–46, which are the first three β strands of GFP, with themselves; they have little inter-residue fluctuation (very little blue). Quadrant 2 represents the inter-residue fluctuations between β_{1-3} and β_{4-6} , and shows that the two groups move in a similar manner and do not have great fluctuations between them. Quadrant 3 shows the inter-residue fluctuations between β_{1-3} and β_{7-11} , more blue is visible indicating little correlation and large fluctuations between the two groups of beta sheets. Similar behavior is observed in quadrant 4, which compares β_{7-11} and β_{4-6} . Quadrant 5 depicts the relatively large fluctuations of β_{7-11} , which has an average fluctuation of 1.25 Å even after the extremely mobile residues 209–211 have been removed from the count. That is nearly four times the average fluctuation of β_{1-6} , 0.32 Å, and is higher than either of the two other groups of β sheets (β_{1-3} , β_{4-6}) with correlated movement shown in Fig. 2. These β sheets have average fluctuations of 0.39 Å and 0.85 Å respectively. Similar behavior was observed in our ANM simulations of other FPs (Figs. S4, S5), in our MD simulations where β_{7-11} are more mobile than the other beta-sheets, and it has precedent in the literature where the last five sheets making up the beta-barrel are predicted to unfold first (56).

Conclusion

Three very different techniques have been used to explore the structures of the mature and immature forms of all naturally occurring GFP-like proteins in the pdb and although the methods are very different the results are surprisingly similar. The Anisotropic Network Model analysis of all the naturally occurring FPs and their structural variation, as well as the molecular dynamics simulations of avGFP revealed that

- Conserved lid residues have lower RMSDs than non-conserved lid residues and have greater dihedral freedom in the MD simulations suggesting that the conserved lid residues might be acting as hinge and/or folding nuclei.
- The mature forms of the fluorescent proteins are much more rigid than the immature structures. Presumably this is due to the fact that upon chromophore formation the β -barrel residues form hydrogen bonds with the chromophore and the barrel constricts to form a rigid structure.
- Sheets β_{1-3} , β_{4-6} and β_{7-11} move as groups, with most of the mobility observed in sheets β_{7-11} , which may reflect GFP's tendency to unfold from β_{7-11} .

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by an allocation of advanced computing resources provided by the National Science Foundation. The computations were performed on Kraken at the National Institute for Computational Sciences (<http://www.nics.tennessee.edu/>). MZ is the Jean Tempel '65 Professor of Chemistry. The research was funded by the NIH (Area Grant # R15 GM59108-04). NSF-STEM 0726589 for RS and PP.

References

1. Ong WJ, Alvarez S, Leroux IE, Shahid RS, Samma AA, Peshkepija P, Morgan AL, Mulcahy S, Zimmer M. Function and structure of GFP-like proteins in the protein data bank. *Mol Biosyst.* 2011; 7:984–992. [PubMed: 21298165]
2. Shagin DA, Barsova EV, Yanushevich YG, Fradkov AF, Lukyanov KA, Labas YA, Semenova TN, Ugalde JA, Meyers A, Nunez JM, Widder EA, Lukyanov SA, Matz MV. GFP-like proteins as ubiquitous metazoan superfamily: Evolution of functional features and structural complexity. *Molecular Biology and Evolution.* 2004; 21:841–850. [PubMed: 14963095]
3. Alieva NO, K.A. K, Field SF, Meleshkevitch EA, Beltran-Ramirez V, Miller DJ, Salih A, Wiedenmann J, Matz MV. Diversity and Evolution of Coral Fluorescent Proteins. *Plos One.* 2008; 3:e2680. [PubMed: 18648549]
4. Stepanenko OV, Shcherbakova DM, Kuznetsova IM, Turoverov KK, Verkhusha VV. Modern fluorescent proteins: from chromophore formation to novel intracellular applications. *Biotechniques.* 2011; 51:313–+. [PubMed: 22054544]
5. Zimmer M. GFP: from jellyfish to the Nobel prize and beyond. *Chemical Society Reviews.* 2009; 38:2823–2832. [PubMed: 19771329]
6. Dedecker P, De Schryver FC, Hofkens J. Fluorescent Proteins: Shine on, You Crazy Diamond. *Journal of the American Chemical Society.* 2013; 135:2387–2402. [PubMed: 23317378]
7. Hoffman, RM. Live cell imaging in live animals with fluorescent proteins. In: Conn, PM., editor. *Imaging and Spectroscopic Analysis of Living Cells: Imaging Live Cells in Health and Disease.* 2012. p. 197-224.
8. Chudakov DM, Matz MV, Lukyanov S, Lukyanov KA. Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiological Reviews.* 2010; 90:1103–1163. [PubMed: 20664080]
9. Onuchic JN, LutheySchulten Z, Wolynes PG. Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem.* 1997; 48:545–600. [PubMed: 9348663]
10. Onuchic JN, Wolynes PG. Theory of protein folding. *Current Opinion in Structural Biology.* 2004; 14:70–75. [PubMed: 15102452]
11. Liu Y, Bahar I. Sequence Evolution Correlates with Structural Dynamics. *Molecular Biology and Evolution.* 2012; 29:2253–2263. [PubMed: 22427707]
12. Tang GW, Altman RB. Remote Thioredoxin Recognition Using Evolutionary Conservation and Structural Dynamics. *Structure.* 2011; 19:461–470. [PubMed: 21481770]
13. Dill KA, Chan HS. From Levinthal to pathways to funnels. *Nature Structural Biology.* 1997; 4:10–19.
14. Mirny LA, Shakhnovich EI. Universally conserved positions in protein folds: Reading evolutionary signals about stability, folding kinetics and function, *Journal of Molecular Biology.* 1999; 291:177–196.
15. Temiz NA, Bahar I. Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase. *Proteins-Structure Function and Genetics.* 2002; 49:61–70.
16. Crameri A, Whitehorn EA, Tate E, Stemmer WPC. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* 1996; 14:315–319. [PubMed: 9630892]

17. Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autooxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA*. 1994; 91:12501–12504. [PubMed: 7809066]
18. Lemay NP, Morgan AL, Archer EJ, Dickson LA, Megley CM, Zimmer M. The role of the tight-turn, broken hydrogen bonding, Glu222 and Arg96 in the posttranslational green fluorescent protein chromophore formation. *Chemical Physics*. 2008; 348:152–160. [PubMed: 19079566]
19. Capraro DT, Roy M, Onuchic JN, Gosavi S, Jennings PA. beta-Bulge triggers route-switching on the functional landscape of interleukin-1 beta. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:1490–1493. [PubMed: 22307602]
20. Andrews BT, Capraro DT, Sukowska JI, Onuchic JN, Jennings PA. Hysteresis as a Marker for Complex, Overlapping Landscapes in Proteins. *Journal of Physical Chemistry Letters*. 2013; 4:180–188. [PubMed: 23525263]
21. Andrews BT, Roy M, Jennings PA. Chromophore Packing Leads to Hysteresis in GFP. *Journal of Molecular Biology*. 2009; 392:218–227. [PubMed: 19577576]
22. Li BS, Shahid R, Peshkepija P, Zimmer M. Water diffusion in and out of the beta-barrel of GFP and the fast maturing fluorescent protein, TurboGFP. *Chemical Physics*. 2012; 392:143–148. [PubMed: 22582003]
23. Andrews BT, Schoenfish AR, Roy M, Waldo G, Jennings PA. The rough energy landscape of superfolder GFP is linked to the chromophore. *Journal of Molecular Biology*. 2007; 373:476–490. [PubMed: 17822714]
24. Orte A, Craggs TD, White SS, Jackson SE, Klenerman D. Evidence of an intermediate and parallel pathways in protein unfolding from single-molecule fluorescence. *Journal of the American Chemical Society*. 2008; 130:7898–7907. [PubMed: 18507381]
25. Mickler M, Dima RI, Dietz H, Hyeon C, Thirumalai D, Rief M. Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104:20268–20273. [PubMed: 18079292]
26. Eyal E, Bahar I. Toward a molecular understanding of the anisotropic response of proteins to external forces: Insights from elastic network models. *Biophysical Journal*. 2008; 94:3424–3435. [PubMed: 18223005]
27. Huang JR, Craggs TD, Christodoulou J, Jackson SE. Stable intermediate states and high energy barriers in the unfolding of GFP. *Journal of Molecular Biology*. 2007; 370:356–371. [PubMed: 17512539]
28. Enoki S, Maki K, Inobe T, Takahashi K, Kamagata K, Oroguchi T, Nakatani H, Tomoyori K, Kuwajima K. The equilibrium unfolding intermediate observed at pH 4 and its relationship with the kinetic folding intermediates in green fluorescent protein. *Journal of Molecular Biology*. 2006; 361:969–982. [PubMed: 16889795]
29. Reddy G, Liu ZX, Thirumalai D. Denaturant-dependent folding of GFP. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:17832–17838. [PubMed: 22778437]
30. Craggs TD. Green fluorescent protein: structure, folding and chromophore maturation. *Chemical Society Reviews*. 2009; 38:2865–2875. [PubMed: 19771333]
31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Databank. *Nucleic Acids Research*. 2000; 28:235–242. [PubMed: 10592235]
32. Yang F, Moss L, Phillips G. The molecular structure of green fluorescent protein. *Nature Biotech*. 1996; 14:1246–1251.
33. Prescott M, Ling M, Beddoe T, Oakley AJ, Dove S, Hoegh-Guldberg O, Devenish RJ, Rossjohn J. The 2.2 Å crystal structure of a pocilloporin pigment reveals a nonplanar chromophore conformation. *Structure*. 2003; 11:275–284. [PubMed: 12623015]
34. Petersen J, Wilmann PG, Beddoe T, Oakley AJ, Devenish RJ, Prescott M, Rossjohn J. The 2.0-angstrom crystal structure of eqFP611, a far red fluorescent protein from the sea anemone *Entacmaea quadricolor*. *Journal of Biological Chemistry*. 2003; 278:44626–44631. [PubMed: 12909624]

35. Tsutsui H, Karasawa S, Shimizu H, Nukina N, Miyawaki A. Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. *Embo Reports*. 2005; 6:233–238. [PubMed: 15731765]
36. Wilmann PG, Petersen J, Pettikiriachchi A, Buckle AM, Smith SC, Olsen S, Perugini MA, Devenish RJ, Prescott M, Rossjohn J. The 2.1 angstrom crystal structure of the far-red fluorescent protein HcRed: Inherent conformational flexibility of the chromophore. *Journal of Molecular Biology*. 2005; 349:223–237. [PubMed: 15876379]
37. Tubbs JL, Tainer JA, Getzoff ED. Crystallographic structures of Discosoma red fluorescent protein with immature and mature chromophores: Linking peptide bond trans-cis isomerization and acylimine formation in chromophore maturation. *Biochemistry*. 2005; 44:9833–9840. [PubMed: 16026155]
38. Nienhaus K, Nienhaus GU, Wiedenmann J, Nar H. Structural basis for photoinduced protein cleavage and green-to-red conversion of fluorescent protein EosFP. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:9156–9159. [PubMed: 15964985]
39. Henderson JN, Remington SJ. Crystal structures and mutational analysis of amFP486, a cyan fluorescent protein from *Anemonia majano*. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:12712–12717. [PubMed: 16120682]
40. Nienhaus K, Renzi F, Vallone B, Wiedenmann J, Nienhaus GU. Chromophore-protein interactions in the anthozoan green fluorescent protein asFP499. *Biophysical Journal*. 2006; 91:4210–4220. [PubMed: 16980366]
41. Nienhaus K, Renzi F, Vallone B, Wiedenmann J, Nienhaus GU. Exploring chromophore-protein interactions in fluorescent protein cmFP512 from *Cerianthus membranaceus*: X-ray structure analysis and optical spectroscopy. *Biochemistry*. 2006; 45:12942–12953. [PubMed: 17059211]
42. Suto K, Masuda H, Takenaka Y, Tsuji FI, Mizuno H. Structural basis for redshifted emission of a GFP-like protein from the marine copepod *Chiridius poppei*. *Genes to Cells*. 2009; 14:727–737. [PubMed: 19469881]
43. Hayashi I, Mizuno H, Tong KI, Furuta T, Tanaka F, Yoshimura M, Miyawaki A, Ikura M. Crystallographic evidence for water-assisted photo-induced peptide cleavage in the stony coral fluorescent protein kaede. *Journal of Molecular Biology*. 2007; 372:918–926. [PubMed: 17692334]
44. Chan MCY, Karasawa S, Mizuno H, Bosanac I, Ho D, Prive GG, Miyawaki A, Ikura M. Structural characterization of a blue chromoprotein and its yellow mutant from the sea anemone *Cnidopus japonicus*. *Journal of Biological Chemistry*. 2006; 281:37813–37819. [PubMed: 17028187]
45. Pletneva N, Pletnev V, Tikhonova T, Pakhomov AA, Popov V, Martynov VI, Wlodawer A, Dauter Z, Pletnev S. Refined crystal structures of red and green fluorescent proteins from the button polyp *Zoanthus*. *Acta Crystallographica Section D-Biological Crystallography*. 2007; 63:1082–1093.
46. Wilmann PG, Turcic K, Battad JM, Wilce MCJ, Devenish RJ, Prescott M, Rossjohn J. The 1.7 angstrom crystal structure of Dronpa: A photoswitchable green fluorescent protein. *Journal of Molecular Biology*. 2006; 364:213–224. [PubMed: 17010376]
47. Pletneva NV, Pletnev SV, Chudakov DM, Tikhonova TV, Popov VO, Martynov VI, Wlodawer A, Dauter Z, Pletnev VZ. Three-dimensional structure of yellow fluorescent protein zYFP538 from *Zoanthus* sp at the resolution 1.8 angstrom. *Russian Journal of Bioorganic Chemistry*. 2007; 33:390–398.
48. Loening AM, Fenn TD, Gambhir SS. Crystal structures of the luciferase and green fluorescent protein from *Renilla reniformis*. *Journal of Molecular Biology*. 2007; 374:1017–1028. [PubMed: 17980388]
49. Violot S, Carpentier P, Blanchoin L, Bourgeois D. Reverse pH-Dependence of Chromophore Protonation Explains the Large Stokes Shift of the Red Fluorescent Protein mKeima. *Journal of the American Chemical Society*. 2009; 131:10356–+. [PubMed: 19722611]
50. Mizuno H, Kumar Mal T, Waelchli M, Kikuchi A, Fukano T, Ando R, Jeyakanthan J, Taka J, Shiro Y, Ikura M, Miyawaki A. Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:9927–9932.

51. Kikuchi A, Fukumura E, Karasawa S, Mizuno H, Miyawaki A, Shiro Y. Structural Characterization of a Thiazoline-Containing Chromophore in an Orange Fluorescent Protein, Monomeric Kusabira Orange. *Biochemistry*. 2008; 47:11573–11580. [PubMed: 18844376]
52. Malo GD, Wang M, Wu D, Stelling AL, Tonge PJ, Wachter RM. Crystal structure and Raman studies of dsFP483, a cyan fluorescent protein from *Discosoma striata*. *J Mol Biol*. 2008; 378:871–886. [PubMed: 18395223]
53. Pletnev S, Gurskaya NG, Pletneva NV, Lukyanov KA, Chudakov DM, Martynov VI, Popov VO, Kovalchuk MV, Wlodawer A, Dauter Z, Pletnev V. Structural Basis for Phototoxicity of the Genetically Encoded Photosensitizer KillerRed. *Journal of Biological Chemistry*. 2009; 284:32028–32039. [PubMed: 19737938]
54. Pletnev S, Morozova KS, Verkhusha VV, Dauter Z. Rotational order/disorder structure of fluorescent protein FP480. *Acta Crystallographica Section D-Biological Crystallography*. 2009; 65:906–912.
55. Ebisawa T, Yamamura A, Kameda Y, Hayakawa K, Nagata K, Tanokura M. The structure of mAG, a monomeric mutant of the green fluorescent protein Azami-Green, reveals the structural basis of its stable green emission. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications*. 2010; 66:485–489.
56. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*. 2013; 27:221–234. [PubMed: 23579614]
57. Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: a software program for pK (a) prediction and protonation state generation for drug-like molecules. *Journal of Computer-Aided Molecular Design*. 2007; 21:681–691. [PubMed: 17899391]
58. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still WC. MacroModel - an Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *Journal of Computational Chemistry*. 1990; 11:440–467.
59. Chang G, Guida WC, Still WC. An internal-coordinate monte carlo method for searching conformational space. *J. Am. Chem. Soc*. 1989; 111:4379–4386.
60. Saunders M, Houk K, Wu Y-D, Still W, Lipton M, Chang G, Guida W. Conformations of cycloheptadecane. A comparison of methods for conformational searching. *J. Am. Chem. Soc*. 1990; 112:1419.
61. Bartol J, Comba P, Melter M, Zimmer M. Conformational searching of transition metal compounds. *J Comput Chem*. 1999; 20:1549–1558.
62. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics*. 2006; 7:339–349. [PubMed: 16836757]
63. Yang F, Moss LG, Phillips GN. The molecular structure of green fluorescent protein. *Nature Biotechnology*. 1996; 14:1246–1251.
64. Kolossvary I, Keseru GM. Hessian-free low-mode conformational search for large-scale protein loop optimization: Application to c-jun N-terminal kinase JNK3. *Journal of Computational Chemistry*. 2001; 22:21–30.
65. Keseru GM, Kolossvary I. Fully flexible low-mode docking: Application to induced fit in HIV integrase. *Journal of the American Chemical Society*. 2001; 123:12708–12709. [PubMed: 11741448]
66. Bowers, KJ.; Chow, E.; Xu, H.; Dror, RO.; Eastwood, MP.; Gregersen, BA.; Klepeis, JL.; Kolossvary, I.; Moraes, MA.; Sacerdoti, FD.; Salmon, JK.; Shan, Y.; Shaw, DE. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing (SC06)*; Tampa, FL: ACM Press; 2006.
67. Berendsen, HJC.; Postma, JPM.; Gunsteren, W. F. v.; Hermans, J. In: Pullman, B., editor. *Intermolecular forces : proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry* held in Jerusalem, Israel; Dordrecht, Holland; Boston, U.S.A: D. Reidel; Apr 13–16. 1981 p. 331
68. Eyal E, Yang LW, Bahar I. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics*. 2006; 22:2619–2627. [PubMed: 16928735]

69. Cantu DC, Forrester MJ, Charov K, Reilly PJ. Acyl carrier protein structural classification and normal mode analysis. *Protein Science*. 2012; 21:655–666. [PubMed: 22374859]
70. Bakan A, Meireles LM, Bahar I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics*. 2011; 27:1575–1577. [PubMed: 21471012]
71. Xu CY, Tobi D, Bahar I. Allosteric changes in protein structure computed by a simple mechanical model: Hemoglobin T <-> R2 transition. *Journal of Molecular Biology*. 2003; 333:153–168. [PubMed: 14516750]
72. Zimmer M. Molecular mechanics, data and conformational analysis of first-row transition metal complexes in the Cambridge Structural Database. *Coordination Chemistry Reviews*. 2001; 212:133–163.
73. Wood TI, Barondeau DP, Hitomi C, Kassmann CJ, Tainer JA, Getzoff ED. Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry*. 2005; 44:16211–16220. [PubMed: 16331981]
74. Barondeau DP, Kassmann CJ, Tainer JA, Getzoff ED. Understanding GFP posttranslational chemistry: Structures of designed variants that achieve backbone fragmentation, hydrolysis, and decarboxylation. *Journal of the American Chemical Society*. 2006; 128:4685–4693. [PubMed: 16594705]

Research Highlights

- Conserved lid residues in FPs may act as hinge residues or folding nuclei.
- Mature forms of FPs are more rigid and compact than immature forms.
- Compaction associated with maturation may aid in chromophore formation.

Figure 1. Anionic (A) and neutral (B) forms of the chromophore. Neutral precyclized (immature) form (C).

Figure 2.

Splay representation of GFP. The β sheet numbering is taken from the PDB, and is used throughout the paper. Conserved residues in the lids are indicated by yellow circles. The β sheets are colored by their unfolding behavior. The β -sheets unfold in groups. In the primary unfolding mechanism the brown C-terminus strands are the first to unfold, while in secondary pathways the green N-terminus strands have been observed to be the first to detach from the β barrel (56). Sheets β_{1-3} , β_{4-6} and β_{7-11} move as groups.

Figure 3.
Side view of the overlap of the 28 naturally occurring GFP-like proteins found in the pdb.
(rainbow colored with red as smallest RMSD and purple as biggest RMSD).

Figure 4.

RMSD of the residues of the phenol, phenolate and immature forms of avGFP averaged over the 50 ns MD simulation. Secondary structure is displayed above the abscissa, as are the most highly conserved residues (>70%, indicated by ■). Lid residues are identified and their location in the bottom (with the C and N termini) or top lid are indicated (— or —, respectively).

Figure 5.

Heat map of the inter-residue distance fluctuation of 1GFL in the first (slowest) vibrational mode. The axes are labeled with the residues. Small fluctuations are colored in red, and larger ones in blue. The regions marked A belong to the kinked alpha helix in the center of the β barrel, and do not fluctuate greatly compared to the other residues. B marks residue 210, the turn between β sheets 10 and 11 and fluctuates the most in the first vibrational mode. The diagonal pattern of red lines visible throughout the image, and partly labeled as C indicate the 11 β strands, which have little to no movement relative to each other.

Table 1

RMDSs of the alpha carbons in 28 wild-types as compared with avGFP after structural alignment, see Table S1 for data.

	Mature RMSD (Å)	Immature RMSD (Å)
All	0.95	1.35
Top Lid	1.04	1.76
Bottom Lid	1.38	1.59
All Conserved	0.84	1.32
Top Lid, Conserved	0.82	1.55
Bottom Lid, Conserved	1.19	1.41

Table 2

The average RMSDs of the backbone atoms from the anionic, neutral and immature avGFP simulations and the standard circular deviation of the dihedral angles from the immature structure. See Table S2 for RMSDs, and Tables S3 and S4 for the phi and psi circular standard deviation, respectively.

	Anionic RMSD (Å)	Neutral RMSD (Å)	Immature RMSD (Å)	Immature Phi Deviation (Degrees)	Immature Psi Deviation (Degrees)
All residues	.86	.84	.93	15.45	15.25
Lid residues	.97	.92	.97	15.94	16.08
Top Lid	.99	.89	.91	16.07	16.29
Bottom Lid	.94	.96	1.03	15.77	15.84
All conserved	.78	.73	.685	15.56	19.36
Top Lid conserved	1.06	.77	.762	19.07	21.74
Bottom Lid conserved	.55	.60	.633	8.52	14.11
$\beta_1 - \beta_6$	.76	.77	.841	--	--
$\beta_7 - \beta_{11}$	.98	.91	1.06	--	--