Connecticut College
Digital Commons @ Connecticut College

Chemistry Honors Papers | Chemistry Department

2021

Why are Gly31, Gly33 and Gly35 highly conserved in all fluorescent proteins?

Christian Salguero
Connecticut College, csalguero05@gmail.com

Follow this and additional works at: https://digitalcommons.conncoll.edu/chemhp

 Part of the Chemistry Commons

Recommended Citation

Salguero, Christian, "Why are Gly31, Gly33 and Gly35 highly conserved in all fluorescent proteins?" (2021). *Chemistry Honors Papers*. 30.
https://digitalcommons.conncoll.edu/chemhp/30

This Honors Paper is brought to you for free and open access by the Chemistry Department at Digital Commons @ Connecticut College. It has been accepted for inclusion in Chemistry Honors Papers by an authorized administrator of Digital Commons @ Connecticut College. For more information, please contact bpancier@conncoll.edu.
The views expressed in this paper are solely those of the author.

"Why are Gly31, Gly33 and Gly35 highly conserved in all fluorescent proteins?"

Christian Salguero

Connecticut College, Department of Chemistry

New London, Connecticut 06320

Spring 2021

Table of Contents

Acknowledgements

I would like to extend my gratitude towards Professor Marc Zimmer for providing the necessary guidance and support for the completion of this thesis and chemistry degree. I would also like to demonstrate my gratitude towards Professor Stanton Ching, who with Marc, jump started my resume by allowing me to conduct research in their labs. A huge thanks to the chemistry department for dealing with both Justin and I goofing around in their classes and labs, while also providing us with the support to succeed. The program Science Leaders has supported both of us continuously and provided the network that has allowed us to pursue a number of opportunities. I would also like to thank Professor Bruce Branchini and Stanton Ching, both of whom have read my honor thesis and provided the necessary feedback. Most of all I would like to thank Justin Nwafor for being a great lab partner, roommate and friend throughout college and the completion of this thesis, love you like a brother and will miss the arguments and our side job as students. To everyone who has worked in or is currently working in the Zimmer lab including: Franceine Welcome, Sercan Durmus, Sophia Moroney and Admirabilis Kalolella, thanks for the good laughs and vibes. The final and biggest thank you goes to my family, all my friends, and the big man himself, God, that has heard my struggles and supported me unconditionally (They really be knowing how low my motivation got during these crazy four years). All in all, grateful for these unforgettable four years.

Abstract

Green Fluorescent Protein (GFP) has grown in popularity and new applications are currently being developed. There are certain residues that are highly conserved through all the naturally occurring fluorescent proteins variants, and some of their functionality is yet to be determined. This is the case for three glycines that appear in a GXGXGX motif in the second β-sheet at positions 31, 33, and 35 with conservations of 100%, 87% and 95%, respectively.

Molecular dynamic simulations and other computational analyses of G31A, G33A, and G35A mutants, derived from pre-cyclized wild-type GFP, determined with confidence that these glycines are not involved in the chromophore formation. It is now suspected that these glycines contribute to the folding pathways of the β-barrel, due to their innate flexibility and small size. Key distances within the structures were measured such as the hydrogen bond network of the α-helix and tight turn to possibly determine the glycines functionality. Quantification of water channels within the protein was completed for all the mutants in order to determine the water migration pathways within the β-barrel. It was determined that the number of overall channels increased, but those with the directionality towards the α-helical region decreased.

Other mutants of the precyclized wild-type GFP included G35V, F71L, F71Y, and G35V/F71L. These mutations were aimed to explore the steric effects and aromatic rescue interactions between the glycines and their neighbouring strands (β1-3). It was determined that there was an increase in the distances within the H-bond network of these mutants, decreasing the rigidity of the β-barrel. The biggest increase was seen in G35V/F71L, due to steric effects, and F71Y, possibly due to steric effects and the charge character that was added.

Introduction

History of FPs:

Bioluminescence has been observed in different aquatic and terrestrial organisms such as jellyfish, fireflies, sea cacti, and squids in different epochs.[1] The organisms that exhibit this trait have a wide variety of utilities for it such as communication, defense and even predatory usages.[1] The bioluminescent process involves an enzyme and a substrate that chemically react with each other in order to produce the observable light within the organism. The enzyme is typically a form of luciferase and the substrate that binds to it is a form of luciferin.[1,2] Some species that possess bioluminescence also express fluorescent proteins (FPs). Unlike bioluminescence, fluorescence requires absorption of a certain wavelength of light from an external source to almost immediately output a lower energy wavelength. The most famous example of this is the jellyfish known as *Aequorea victoria*.[1,2] Its bioluminescent reaction produces a blue light, but it's fluorescent protein converts it to an observable green. Not all fluorescent organisms have bioluminescence; for example, reef corals and some species of shrimps are fluorescent but lack bioluminescence. Some researchers believe that the evolution of fluorescent proteins are part of environmental adaptations of species.[3] For example, researchers believe that the fluorescent proteins present in coral reefs are an adaptation to protect them from the constant exposure to ultraviolet light in the shallow waters.[3,15] Other possible uses that these proteins have are that they can serve as primitive proton pumps or light-induced e^- donors.[4,15]

Green fluorescent protein (GFP) is a protein that has become widely known and used in recent years. GFP was first discovered in the jellyfish, *Aequorea aequorea,* or more commonly known as *Aequorea victoria.*[1] The first reported *Aequorea* fluorescence was in 1955 when the jellyfish was irradiated with an ultraviolet light.[1,2,4] The jellyfish bioluminescence involves two proteins, aequorin and GFP.[1,2] The *Aequorea victoria* GFP was the first fluorescent protein to be isolated, clone, and used in a variety of experiments as a tracer.[1,2] This protein takes the blue light generated from the reaction of three

calcium ions in the photoprotein aequorin and coelenterazine, and radiationless converts it into green light.[1,2] The first crystallization of GFP occurred in 1974 and diffraction patterns were analyzed and reported in 1988.[5] Finally by 1996, the structure of GFP was solved and uploaded to the Protein Data Bank by Ormo and Phillips.[2,5]

Osamu Shimomura isolated a small peptide fragment that contained the chromophore from a heat-denatured GFP.[1] From this Shimomura was able to synthesize a handful of compounds which he compared to the GFP's chromophore.[1] By 1979, he was able to propose the chemical structure, 4-(*p*-hydroxybenzylidene)imidazoilid-5-one, the first of its kind.[1] Further research confirmed Shimomura's proposed structure and that it was a cyclic hexapeptide formed by residues 64-69 (Fig 1).[1]

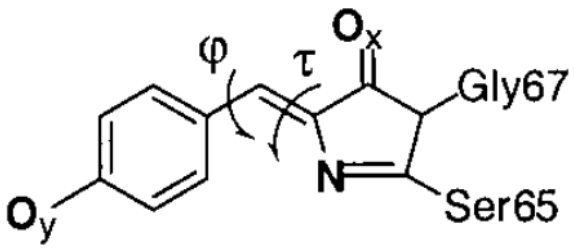


Figure 1. The 4-(*p*-hydroxybenzylidene)imidazoilid-5-one structure that was proposed by Shimomura in 1979.[1]

GFP has a rare behavior that allows it to attack its own backbone in order to form the fluorescent chromophore, or fluorophore.[1] This behavior was confirmed in 1994 which proved that the chromophore formation was through an intramolecular autocatalytic cyclization (Fig 2).[1]

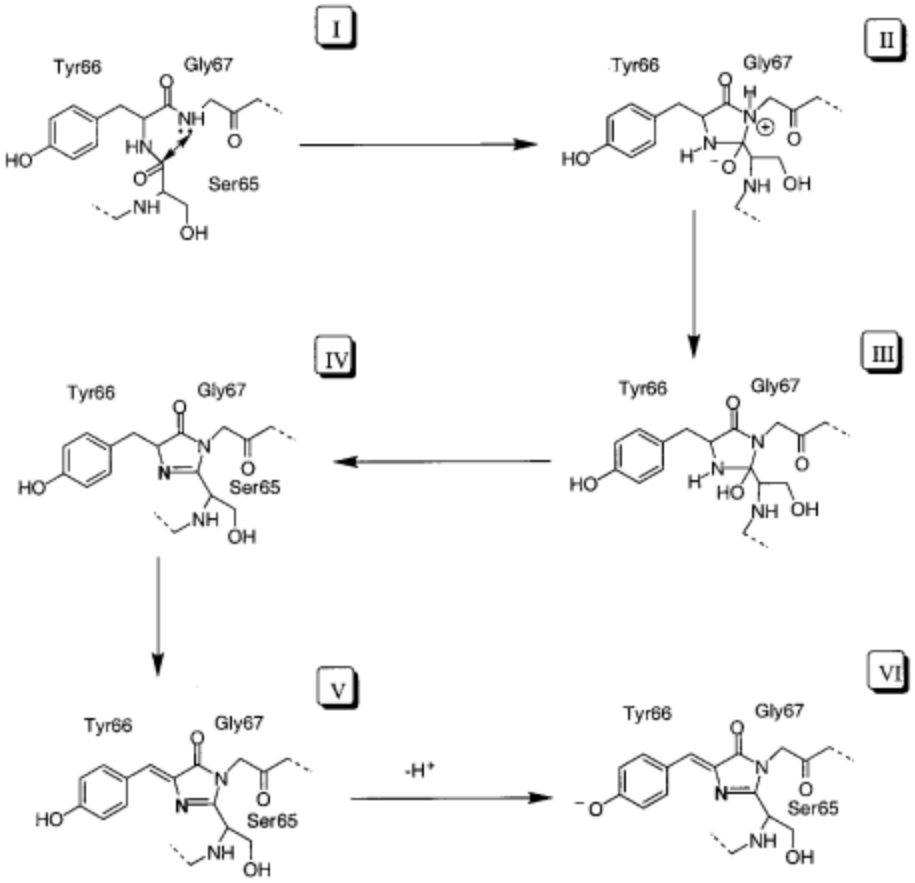


Figure 2. Proposed autocatalytic mechanism for the formation of the chromophore.[1]

The presence and high stability of the chromophore has made it extremely useful as an imaging tool. Many mutations have been performed on GFP in order to widen its use, some of which have changed its color, brightness, and sensing capabilities. The first experimental application of GFP was seen as a gene-detector on the nematode *Caenorhabditis elegans*.[5] Researchers wanted to detect the *C.elegans* gene expression of the *mec-7* promoter in an *in-vivo* environment, this was made possible after the implementation of a GFP tag.[5] Research continues to advance on this protein in order to widen and improve its understanding and usage.

FP Applications:

FPs have become abundant due to their wide range of usages such as photolabeling, tracking, and imaging; all of which have allowed for significant improvements and advancements in a number of fields.

GFP is commonly used due to the simplicity of its formation and the fact that it does not alter the mobility and performance of tagged proteins.[1] GFP is extremely durable, however, environmental factors such as alkaline pH, detergents, photobleaching, and salts can affect is performance.[12] Temperature has been observed to have an inverse relationship to GFP's folding rate.[5] GFP synthesized at lower temperatures has been observed to be stable and fluorescent up to temperatures of 65 ℃, a stability that is lost as the ambient temperature increases.[5] The malleability of this protein has also made it attractive in fields such as cellular research.

Fluorescent proteins allow researchers to track the growth of cells, infections, and the effects that mutations have on an organism by serving as fusion tags.[1] This is done through the creation of a chimera, which fuses the interested protein to the amino or carboxyl termini of the GFP.[1] Expression levels of a targeted gene can be monitored with GFP through its functionality as a reporter gene as seen in the *C. elegans* studies. GFP can also function as an indicator for environmental factors such as metal ions, pH, and halides.[7]

Different codons can be changed in order to optimize it for different environments such as mammals, plants, fungi, etc..[1] For example, changing the bulky groups in the protein to smaller ones can increase the folding efficiency that it has, making it suitable for higher temperatures.[5] Some have even made a variant called "super folder" GFP which can fold in extremely poor conditions, have an improved resistance to chemical denaturants, and increased folding kinetics.[6] The protein's popularity has also increased in recent years due to the different colored analogs of GFP that can be engineered with similar ease (Fig. 3).[3]

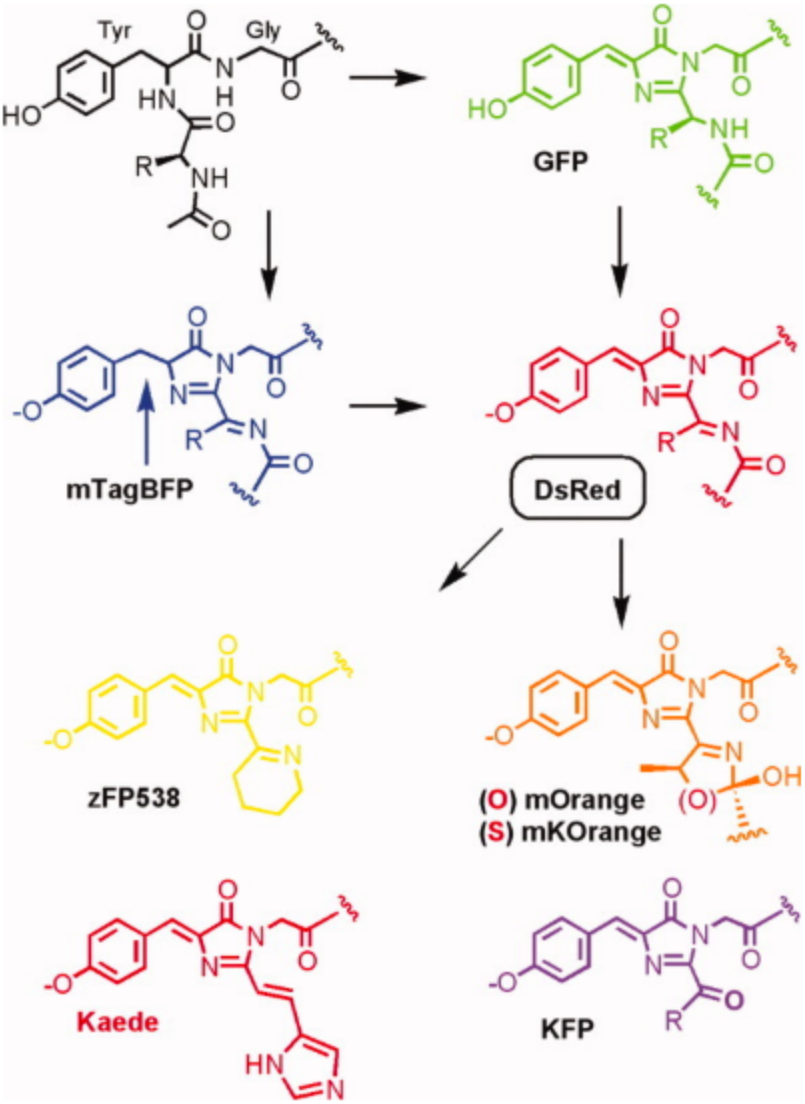


Figure 3. Different color fluorescent proteins that have been discovered, all of which has a λ_{max} = 550 - 650 nm. DsRed (red), mTagBFP (blue), zFP538 (yellow), (O) mOrange (orange), (S) mKOrange (orange), Kaede (green, after photoconversion emissions shifts to red), and KFP (photoswitchable, orange). [7]

The analogs shown in Fig. 3 have similar folding pathways as the wild-type GFP.[3] The structural and sequential differences between these variants are, in some cases, small.[7] Other analogs have been engineered to fluoresce near the infrared spectrum.[8,9] Some of these variations are exceptional bright in mammalian cells and allow for multicolor imaging, making them highly suitable for long term *in vivo*

imaging.[8,9] Their non-invasiveness makes them ideal for research in fields such as cancer studies, stem cell biology, and neuroscience.[8,9]

Limitations:

A prominent field where GFP is used is cancer research. In these studies, GFP allows tracking of tumor progression and possible metastasis. A major drawback to GFP is that it can not become fluorescent in an anaerobic environment, which cancer/tumor cells are known to be, due to the dependence of the key Tyr66-dehydrogenation step on O_2 (Fig 2,step 2).[5] This limitation is eliminated once the GFP has matured. Once maturation has occurred, the presence of O_2 has no effect on its functionality.[5] Upon exhaustion of O_2, a variety of GFP species can undergo a photo conversion into a red fluorescent species.[5]

Even though GFP is relatively large for a tag, there have been numerous cases where the targeted proteins were successfully tagged without losing functionality.[10] Researchers can use three-color and four-color flow cytometry techniques to track multiple variables at the same time in an *in-vivo* environment.[1] Expression of GFP has been successfully achieved in a number of species such as *E. coli, Drosophila, C. elegans,* and yeast, but the expression efficiency and intensity decreases in more complex organism such as mammals and plants.[11] Through other modifications, researchers have been able to use GFP as a convenient reporter for gene regulation, signal transduction, and subcellular localization of chimeric proteins in plants.[11] The only limitation is the fact that each GFP structure has only one chromophore meaning that a high level of expression is needed for proper and accurate visualization.[5,7,11]

FPs can serve as a sensor for specific environmental factors such as pH level, Ca^+ concentration and intercellular processes.[7] For example they can serve as a pH-sensor by switching the color of emission from a blue light (~460 nm) in a low pH environment to green (~510 nm) at a higher pH.[7] However, this functionality is not used commonly or as efficiently as others since researchers are still exploring this application.[7] GFP's 3D structure is highly interesting and thus the subject of our research.

Fluorescent Proteins Structures:

GFPs consist of 238 amino acids including the chromophore forming residues and share the common cylindrical shape formed by 11 β-sheets, each 3 to 10 residues long, with other naturally occurring FPs.[10] This twisted-cylinder formation is called the β-barrel or β-can, which is then threaded by an α-helix, another secondary structure, that is responsible for chromophore formation (Fig. 4,5).[5,16]

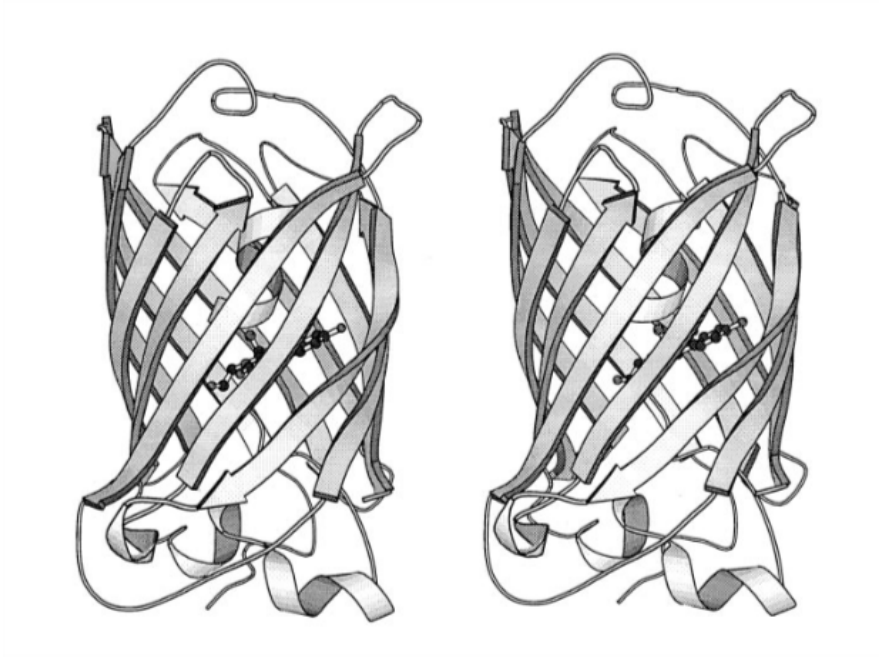


Figure 4. The 3 dimensional structure representation containing the 11 β-strands, α-helix, and chromophore.[5]

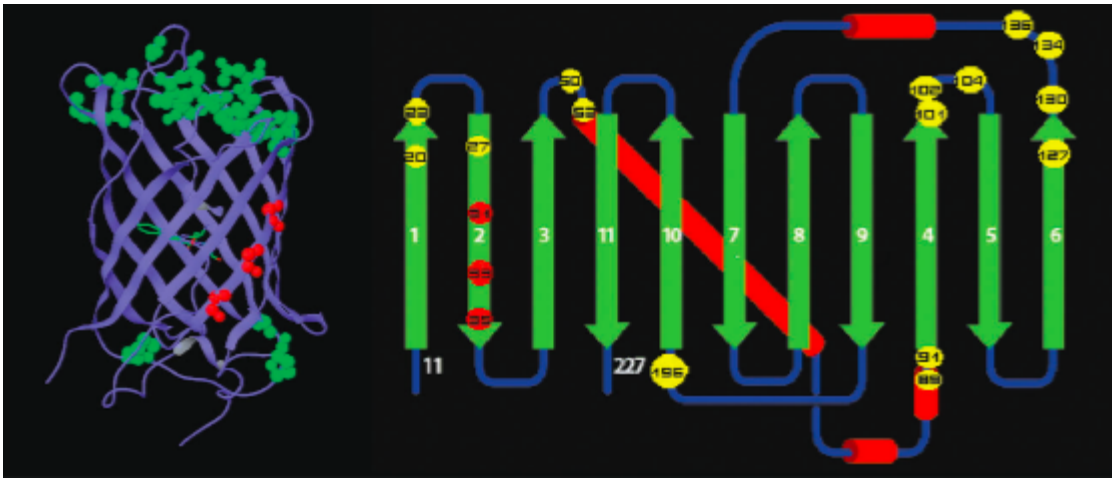


Figure 5. GFP structure both 3D and 2D, with proper labeling of the strands and the 3 highly conserved glycines.[15]

The β-can forms rapidly, but fluorescence does not appear until 90 mins - 4 hours later due to the chromophore's slower cyclization reaction.[1,4] Some engineered GFP variations have minimized this delay through deletions of certain residues that do not affect their overall structure and functionality. GFP's cyclization time is unable to shorten since most of the primary sequence is used in the formation of the β-barrel and the α-helix, thus making deletion hardly possible.[5]

Structural studies have determined the amount of mutations that GFP can endure before fluorescence extinction occurs. One study found that 75% of single mutations have a diminishing effect on fluorescence and another 9.4% of these decreased fluorescence by >5-fold.[17] However, for many of these mutations the effects were small.[17] It was observed that more mutations led to decreased fluorescence and once there were more than five mutations, fluorescence was extinguished.[17]

Other research has indicated that the histidine ammonia lyase (HAL) and phenylalanine ammonia lyase (PAL) protein families have the same post transitional ring formation.[1] The fluorescent chromophore is at the center of the β-barrel in the α-helix where it is protected and formed through the autocyclization of residues Ser65, Tyr66, and Gly67(Figure 1 & 2).[1,2,5,12,13,15 Figure 1,2]

The chromophore can exist in 2 states, neutral and anionic, which affect the wavelength that they absorb.[7] The neutral, or A form, absorbs light around 395 nm, while the anionic, or B form, absorbs a wavelength of 475 nm (Fig. 6).[7]

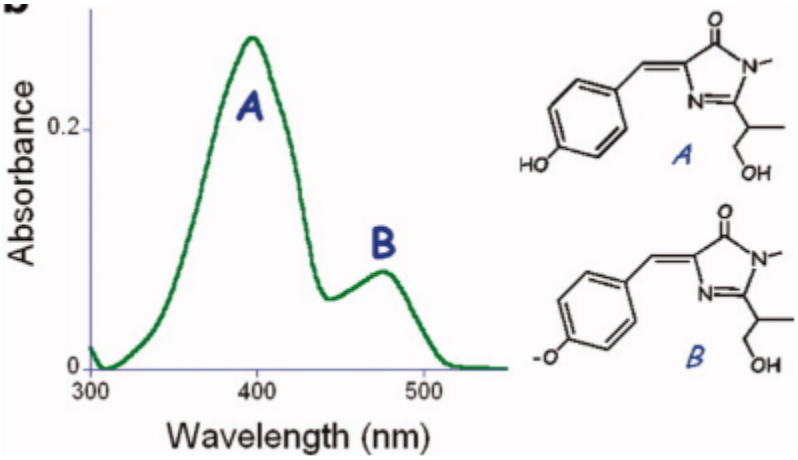


Figure 6. The different states of the chromophore, both the A (neutral) and B(anion) and their perspective peaks.[7]

Both forms of the chromophore exist in the wild-type GFP in a 6:1 ratio, favoring the formation of the A form, and are unaffected by environmental factors such as salinity or pH.[7] The position and the size of the cavity that surrounds the chromophore, which depends on hydrogens bonds and the hydrophobic chains, is vital in determining the fluorescence level.[7]

All naturally occurring chromophores are created by 2 rings that are linked by a double bond.[7] In the chromophore's excited state the two rings are perpendicular to each other, causing the energy to be radiated thermally rather than radiatively.[7] Stereochemistry does not affect its fluorescence, but there is a favorability for the *cis* isomer.[7] This is true for all GFP variants, but in other FPs such as eqFP611, a red fluorescent protein, the *trans* isomer is favored.[7]

Conserved Residues in GFP:

An isolated chromophore is nonfluorescent, either as a naked molecule or as part of the isolated hexapeptide.[7] It is hypothesized that the β-barrel formation is crucial in restraining the chromophore in the proper position for fluorescence promotion.[7] An important residue in chromophore maturation is

Gly67, as its unusual nucleophilic nature promotes the auto-cyclization reaction with minimal steric hindrance.[5] Other conserved residues include Arg96 and Glu222, which were determined to have a catalytic role in chromophore formation.[7] Arg96 was observed to promote ring closure in the chromophore, and Glu222 seems to contribute to the rigidity of the chromophore within the structure, preventing it from moving and becoming nonradiative.[3,15,12] The preservation of Glu222 supports the suggested role of fluorescent proteins as proton pumps.[15] The ground state of the chromophore is deprotonated by Glu222, which is then protonated through a well-designed proton entry pathway.[15] However, experimental data demonstrates that the role of Glu222 is not to contribute to the excited state proton transfer (ESPT) reaction.[15] There are one hundred and fifty one GFP-like structures that conserve Glu222, of them only 56 are in a proton chain.[15] Thus, it is believed that the more important role of Glu222 is in the chromophore maturation rather than the ESPT reactions.[15]

The GFP chromophore is formed through the autocyclization of the residues 65-67 (Ser-Tyr-Gly), a sequence that is not conserved in all fluorescent proteins.[3] Only a fraction of this is conserved in GFP-like proteins: residues 66 and 67 (Tyr-Gly).[3,15] Ser65 can mutate to Lys, Asn, and Gln and still have functionality.[3] However, in GFP, mutations in position 65 cause a significant drop in fluorescence.[3,12] Substitution of Gly67 by any other amino acid will impair the formation of the chromophore and will render GFP non-fluorescent.[12] It is believed that glycine is the only amino acid that will allow the kinked-conformation required for the auto-cyclization reaction. [12] All naturally occurring fluorescent proteins conserved Tyr66, which can be mutated artificially to other aromatic residues (His/Trp) to produce different colored FPs.[12,15] Chromophore maturation can occur with non-aromatic residues at position 66, but side reactions (hydrolysis/fragmentation) and small conjugation can quench its fluorescence.[12,5]

Species of origin	# of structures in PDB	PDB code of wild type	Residues 65, 66, 67, 148, 203, 204, 205, 222
Aequorea victoria	144	1GFL	SYGHTQSE
Discosoma sp.	20	1G7K	QYGTSKLR
Artificial gene	12	—	—
Entacmaea quadricolor	11	1UIS 3H1O	MYGTHRLH, MYGTYRLV
Anemonia sulcata	10	1XMZ	MYGTHRIA
Zoanthus sp.	8	1XAE	KYGCHKLH
Echinophyllia sp.	8	2POX	CYGTHHIH
Lobophyllia hemprichii	5	1ZUX	HYGTHCIH
Anthomedusae	5	Wild type not crystallized[a]	QYGEIITE
Fungia concinna	4	Wild type not crystallized[a]	CYGTHRLD
Montipora sp.	4	Wild type not crystallized[a]	QYGTRKLI
Montipora efflorescens	4	1MOU	QYGSRKLI
Pectiniidae	3	2Z6X	CYGTHRIH
Aequorea coerulescens	3	3LVA	SYGHTQSG
Pontellina plumata	3	2G6X, 2G6Y	GYGVRRVH, GYGVRRVH
Clavularia sp.	3	2OTB	AYGTHRIS
Anemonia majano	3	2A46	KYGFHRIH
Favia favus	3	1XSS	DYGTHRIH
Chiridius poppei	2	2DD7	GYGVRRVQ
Trachyphyllia geoffroyi	2	2GW3 (green form), 2GW4 (red form)	HYGTHCIH
Cnidopus japonicus	2	2IB5	QYGCHRPS
Heteractis crispa	1	1YZW	ES-TIRLA
Renilla reniformis	1	2RH7	SYGYHRLT
Cerianthus membranaceus	1	2C9J	QYGEHRLY
Dendronephthya sp.	1	2VZX	HYGTHRIH
Galaxea fascicularis	1	3ADF	QYGTHRIN
Clytia gregaria	1	2HPW	SYGHVWTE
Discosoma striata	1	3CGL	QYGTTKLI

[a] Wild type not crystallized. Only a chemically/biotechnologically interesting mutant was crystallized. These mutants were not used in the alignment and the weblogo representation shown in Fig. 3.

Figure 7. Table demonstrates the residues 65,66,67 that are present in a number of species as well as the number of structures and other residues present in said species.[15]

Residues located at the “lids" of β-barrel have been observed to be highly conserved due to their suspected functionality as hinges for β-barrel formation.[4,15] Glycine is one of the most conserved residues in GFP variations, which is believed to be due to their stabilizing effect at certain positions.[14] All natural GFP variants conserve Gly31,Gly33, Gly35 located on the second β-sheet (Fig. 7).[15] These residues are not directly involved in the chromophore maturation and the second β-strand does not differ significantly from the other strands, thus the purpose of these highly conserved residues still remains a mystery.[4, 14] It is suspected that these glycines aid in the formation and stability of the protein, but no evidence has been proven. Studies have shown that the residues that have a greater impact on fluorescence are the ones that were oriented internally towards the chromophore due to the influence that these have on the shape of the cavity and chromophore positionality.[17] These conservations have allowed researchers to replicate GFP-like structures from de-novo experimentation.[16] The denovo FP has less than half of the residues that

are present in naturally occurring GFP, a size differentiation that can provide insight on the functionality of highly conserved residues such as the glycines.[15,16]

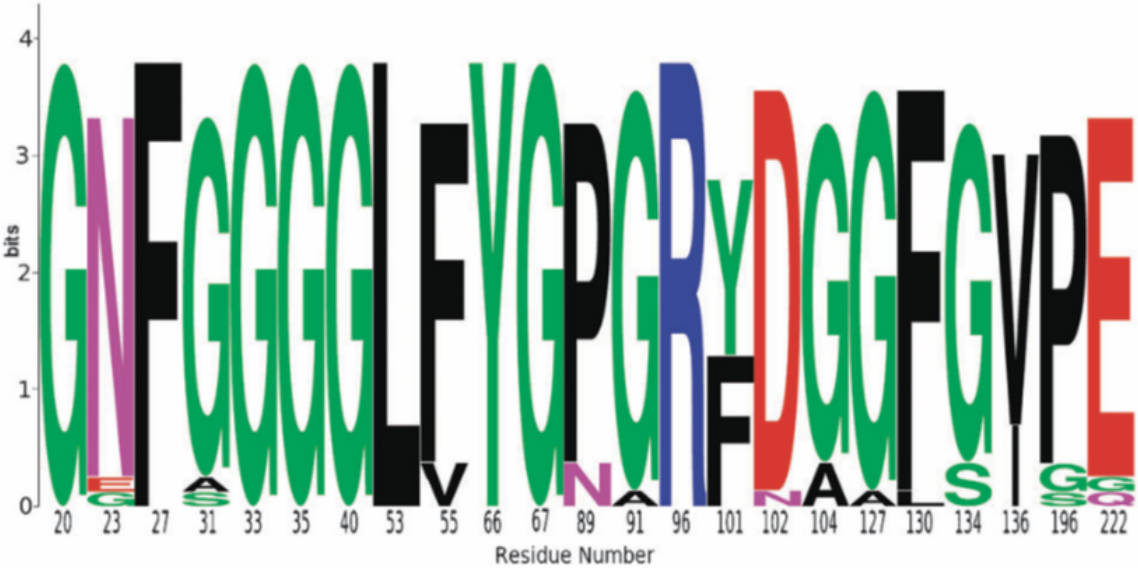


Figure 7. Visualization of the conserved residues through the variations of GFP structures.[15]

The folding pathways of GFP has been studied through unfolding and refolding of the protein in an attempt to understand the conservations.[12] It was determined that the proline residues play an important role in the maturation of the GFP.[12] All the proline residues are in the *trans* form, with the exception of Pro89.[12] The position of this *cis* proline dramatically changes the direction of the chain, thus playing an important role in packing the α helix within the β barrel.[12] The other prolines are suspected to kink the helix into the right conformation.[12] The chromophore seems to play an important role in the overall protein stability after its maturation.[12]

Folding Pathways:

The folding pathway is key in promoting fluorescence in FPs.[18] Computational analyses using multicanonical (MC) molecular dynamics simulations of GFP have shown intermediates that serve as potential energy traps, which prevent proper β-barrel folding and quench fluorescence (Fig. 8).[18] GFP was

unfolded through chemical and mechanical means.[18] Mechanical force was applied to specific residues within the structure.[18] GFP unfolded in either a all-or-none process or by populating nonfluorescent intermediate states, and was refolded through the reduction of the applied force.[18] A commonality between the pathways was that β strands remained grouped in the intermediates such as β-strands 1-6 and the N-terminus β strands.[18] These intermediates provide insights on the pathways that GFP takes to reach its native state and the functionality of specific areas and residues within the structure.[18] It also demonstrated the stabilizing effect that the chromophore maturation provides, since a high energy intermediate was populated in the simulations contrasting the observed stable experimental GFP intermediates.[18]

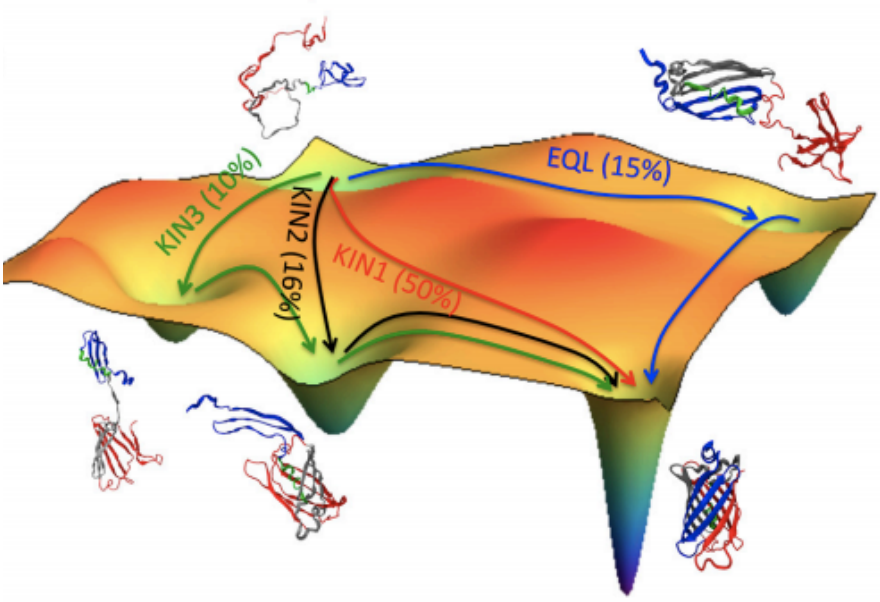


Figure 8. Folding landscape and network of GFP.[18]

FPs Outside of Science:

GFP has mostly been used as a reporter gene in quantitative analysis of either gene expression or other environmental factors. An example of this is the quantification of toxic chemicals within an environment, where the lower fluorescence level indicated higher pollution. The imaging field has also shown an advancement since GFP is non-invasive, allowing the

tagged structure/cell to function normally. Due to its non-invasive behavior, the observation of delicate systems such as brain circuitry and virus infections was made possible both *in vitro* and *in vivo*.[19] However with the development of the new mutants and their different colors, FPs can be used in other fields such as art. Artists either draw inspiration from the structure of the protein or use the protein in their artwork (Fig 9 and 10).[20]

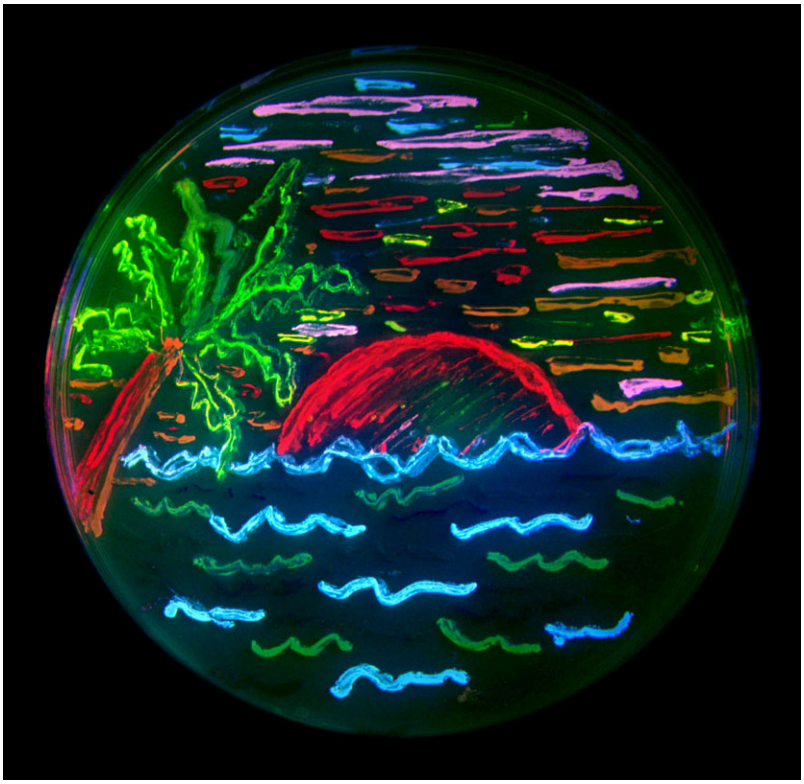


Figure 9. Fluorescent artwork demonstrating the diversity of colors derived from the GFP mutants (credit to Tsien).

Fig 10. Julian Voss-Andreae's sculpture of GFP, "Steel Jellyfish," located at the University of Washington.[20]

Fluorescent pets have also been engineered in order to demonstrate the wide-range of applications that GFP has and demonstrate its non-invasive nature. Alba, the GFP rabbit, was commissioned by Eduardo Kac in a French lab in order to speak to these attributes. However, these animals can also serve the purpose of model-organisms in studies of different diseases such as HIV, narcolepsy, and blindness.[21,22] Other animals that were engineered include: axolotls, zebrafish, cats, beagles, and pigs (Fig 11).[21,22] These animals were commissioned by scientists from all over the world including: the United States, Japan, South Korea, and Taiwan.[21, 22]

Fig 11. Mutate mice (left & right) expressing their fluorescence compared to non-mutated mice (center).[22]

Overall, GFP is a unique protein due to its nature, functionality, and structure. It has been used throughout the world in the improvement of the imaging field, and even allowing the creation of art. Research using GFP will continue to expand at the same rate as our curiosity, eventually becoming an integral part of our understanding of the brain circuitry, cancer, and many other diseases.

Our Research:

We obtain a pre-cyclized GFP structure from the Protein Data Bank (PDB), 2AWJ, which was later graphically mutated to be identical to wild-type GFP (wtGFP). A 200ns molecular dynamic simulation was run to set a baseline for the future experiments. A total of three single point mutants were graphically constructed (G31A, G33A, G35A) to explore the functionality of these conserved residues. Each of the mutants had a 200ns MD simulation with the same parameters, and it was determined that the functionality of these was still unclear. These mutants were synthesized by Professor Schneider, who observed a decrease in fluorescence caused by a large amount of misfolding. Our group suspected that this aggregation to be a result of the G3XA mutations. This led to the belief that the glycines were involved in the protein folding pathway. Moving forward, a set of 4 double mutants was constructed computationally through the mutation G35A and the residue opposite from it (Phe71) in order to explore the interaction between the surrounding residues and the glycine. Other simulations included solely the β-strands to explore the folding pathway of GFP and its β-barrel as suggested by the results of the wet-lab. Strands 1-3 were the focus of these 200ns MD simulations, allowing us to observe their movement and interactions in respect to each other.

Experimental

Baseline and Single-Point Mutants Simulations:

Obtained GFP structures 2AWJ, 2AWK and 1EMB from the Protein Data Bank, all of which have been synthesized and had their structure determined via crystallography. The 2AWJ structure is a precyclized intermediate of 2AWK while 1EMB is a mutant constructed by substituting Gln80 with an arginine.[23] Both 2AWJ and 2AWK have substituted the positively charged arginine at position 96 with a neutral methionine.[24] This mutation slows the chromophore formation from hours to months.[24] Graphically mutated the 2AWJ structure to match the sequence of the immature wtGFP structure (1EMB), these mutations included: L64P, T65S, M96R, S99P, T153M, and A163V.

A 25,000 large-scale low-mode sampling step conformational search was conducted on the wtGFP, the resulting structure was then minimized (10,000 steps) to determine the lowest energy structure to serve as a starting point for MD simulations. Three single-point mutants were constructed graphically from this wtGFP structure by substituting the glycines at positions 31, 33, and 35 with alanines. Another 10,000 step minimization was conducted on each of the mutants. The conformational searches explored different vibrational (large-scale low-mode sampling) and torsional space using the Monte Carlo method in order to determine the lowest possible energy structure.

The program Desmond, a software package designed to run high-speed molecular simulations on existing systems such as GFP and the current mutants, was used to run all of our simulations.[3] The GUI Maestro, a program in the Schrödinger suite, was used to model all the mutants as well as to visualize the data acquired from the MD simulations.[25] The resulting

structures from the MD simulations were overlapped with each other to determine the simulations quality and accuracy.

The Desmond system builder was used to prepare all structures for the MD simulations.[26] Predefined SPC solvent were used in a 10Åx10Åx10Å orthorhombic buffer box around the starting structure.[4] Due to the intrinsic charge that the structures have, 6 Cl^- ions and 0.15 M NaCl, were added for charge neutralization.[26] The engineered system was loaded into the Molecular Dynamic panel for preparations/initialization of the simulations. The temperature, ensemble class, and pressure that were used for all simulations were 300K, NPT, and 1.01325 bar, respectively.[26] Simulation quality and structural equilibrium were confirmed through examination of root-mean-squared deviation (RMSD) measurements.[26] The RMSD value determines the movement that the structure exhibits throughout the simulation. A stable structure will have minimal RMSD fluctuation since it will not be shifting into different conformations. Simulations that had a high RMSD value were extended until an equilibrium was found. The OPLS3 force field was used while conducting our computations.[27,28]

Hydrophobic Pocket Simulations:

Followed the same procedure for the system builder as previously described. These simulations explored the steric interactions between residues Gly35 and Phe71, located across from each other. Engineered a total of 4 mutants from the minimized wtGFP structure, three single (Gly35Val, Phe71Leu and Phe71Tyr) and one double mutant (Gly35Val / Phe71Leu). All had 200ns MD simulations conducted on them.

β-strands Simulations:

β-strands 1-3 were graphically isolated (residues 3-57) from the immature, engineered wild-type structure (wtGFP) to determine folding pathway interactions between them. The

terminus of the structure was capped through the protein preparation wizard tool and minimized (10,000 steps). The three single-point Gly/Ala mutations were made on this minimized structure and another 10,000 minimization was run in preparation for the MD simulations. All of these simulations were 400ns in length.

Results and Discussion

Structural comparison between Mutant & wtGFP and validation:

The structure, 2AWJ was used for these simulations due to this variant having a precyclized form caused by the R96M mutation. This mutation strips position 96 of the positive charge that provides its catalytic property, slowing chromophore formation to months.[24] The slowed chromophore formation allowed for the determination of the pre-cyclized intermediate, 2AWJ, as well as its fully cyclized counterpart, 2AWK. Since the difference between the wild-type GFP and the 2AWJ mutant were the kinetics, this structure once graphically mutated to match the wtGFP sequence, could provide adequate modeling of the pre-cyclized intermediate's behavior. Validation of all the output structures from the molecular dynamic (MD) simulations was determined through observation of RMSD values relative to 1EMB, with a constant value signifying a stable conformation of the structure. The first 100 ns of the mutation G31A simulation was observed to have a constant increase in the RMSD, thus extensions to the simulations were required. Depending on the mutation, the structure found the stable confirmation either earlier or later in the simulation.

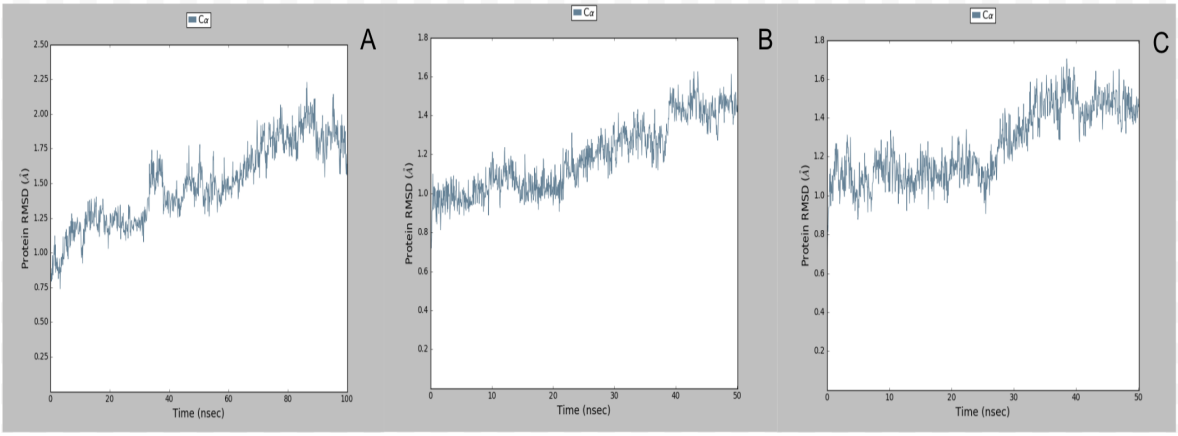


Figure 12. Graphical representation of the G31A mutant's RMSD. From left to right were 0-100ns (A), 100-150ns (B), and 150-200ns (C).

The outputted structure of the engineered 2AWJ was compared, both graphically and quantitatively, with the default 2AWJ, its cyclized counterpart (2AWK), and the wild-type GFP (1EMB) to determine the RMSD values and differences reported in Table 1. Comparison of the RMSD between structures provides an insight into the similarities and differences that the structures possibly have, thus a low RMSD is desired in such studies as our own. The low RMSD values implied that there are only local changes in the structures without any significant global change, thus we can use this as the starting structures for our simulations.

Table 1. RMSD values of the superimpositions and the largest distance difference for the engineered 2AWJ structure.

Structure-engineered 2AWJ	RMSD (Å)	Residues with largest difference, distance
Default 2AWJ	0.8625	G67, 1.734
2AWK	0.5911	Y66, 1.548
1EMB	0.6729	S67, 1.480

The observed low RMSD values indicated that the engineered 2AWJ had a similar structure as the wtGFP and thus it was a suitable pre-cyclized model for the studies. This structure was then modified for all the other mutants: G31A, G33A, G35A, G35V, F71L, F71Y, and G35V/F71L, and validated through the measurement of respective RMSD.

Measurement Acquisitions:

The G3XA mutants were used in order to determine whether the glycine at those positions were contributing to the formation of the chromophore. Using the simulation event analysis built into the Schrödinger suite, different measurements were performed on key distances to draw a comparison between the wtGFP and the mutants. These measurements included: H-bonds interactions/distances between β-sheets, α-helical H-bonds, aromatic rescue interactions, chromophore's "tight turn" distance, and water migration.

Hydrogen distance in helical section of GFP beta barrel:

Hydrogen bonds have a significant effect on the structure of the protein, thus the monitoring of them is essential in determining the effects of certain residues on a protein's overall structural stability.[29] GFP's hydrogen network is unique due to its kinked α-helix structure. This structure has a "sporadic" hydrogen bond pattern, rather than the traditional *i*+4 pattern found in α-helices, thus a comparative analysis of the following distances between the wild type and our mutant, 2AWJ G31A, was performed (Figure 13).[30]

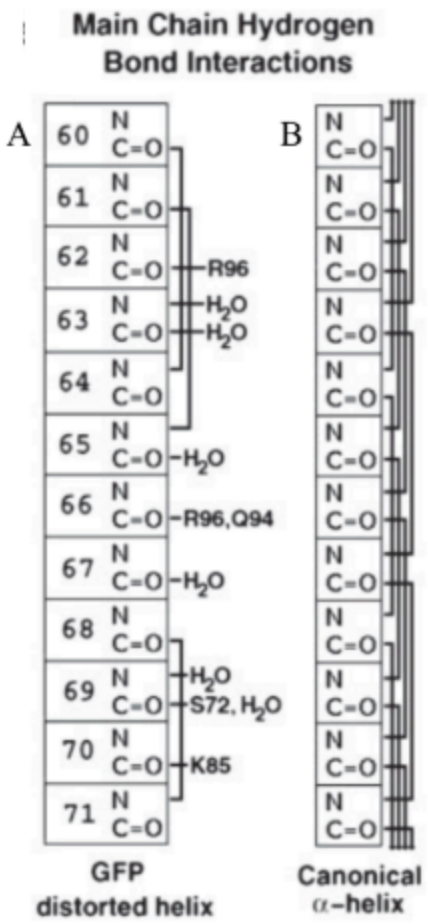


Figure 13. Key distances that were measured during the experiment. The diagram labeled A, demonstrates the H-bonds (solid lines) that are present in the main chain of the alpha helix of GFP. The left diagram B, is a canonical alpha helix and the H-bonds are that formed within it (solid lines). [30]

This unique break in the hydrogen bonding of the α-helix contributes to the formation of the chromophore due to it positioning the reacting residues closer to each other. This reduces the number of H-bonds in the α-helices that need to be broken for the formation of the chromophore.[29] GFP overcomes this energetically unfavorable reaction through its distorted α-helix which disrupts the hydrogen bonding network, allowing the chromophore formation to occur without breaking any H-bonds.[30]

Table 2. Alpha helical, shown in Figure x, H-bond distances in angstrom (Å) for all the wtGFP and all the mutants.

R_1-R_2	60-64	61-65	62-96	66-96	66-94	68-71	69-72	70-85	65-67
2AWJ M96R (wtGFP)									
avg distance (Å)	4.739	3.546	5.411 (N^+) 4.232 (N^0)	8.872 (N^+) 7.519 (N^0)	4.036	3.283	4.318	4.047	3.493
standard deviation	0.327	0.303	0.373 (N^+) 0.362 (N^0)	0.555 (N^+) 0.531 (N^0)	1.089	0.291	0.331	0.447	0.167
G31A (100-150 ns)									
avg distance (Å)	3.638	2.911	4.062 (N^+) 5.145 (N^0)	3.103 (N^+) 4.935 (N^0)	2.895	3.082	3.347	3.563	3.588
standard deviation	0.265	0.159	0.495 (N^+) 0.483 (N^0)	0.326 (N^+) 0.472 (N^0)	0.229	0.222	0.327	0.710	0.190
G31A (150-200 ns)									
avg distance (Å)	4.034	2.936	4.686 (N^+) 4.898 (N^0)	3.010 (N^+) 4.419 (N^0)	3.053	3.090	3.366	5.407	3.353
standard deviation	0.374	0.162	0.668 (N^+) 0.442 (N^0)	0.287 (N^+) 0.531 (N^0)	0.628	0.224	0.367	0.339	0.170
G33A									
avg distance (Å)	4.143	3.897	4.085 (N^+) 3.054 (N^0)	7.898 (N^+) 5.668 (N^0)	6.222	3.400	3.212	3.300	3.295
standard deviation	0.362	0.311	0.398 (N^+) 0.321 (N^0)	0.340 (N^+) 0.363 (N^0)	0.721	0.287	0.229	0.341	0.140
G35A									
avg distance (Å)	4.179	3.000	4.259 (N^+) 5.779 (N^0)	6.697 (N^+) 4.933(N^0)	4.713	3.358	4.218	5.238	3.836
standard deviation	0.315	0.213	0.398 (N^+) 0.456 (N^0)	0.507 (N^+) 0.438 (N^0)	1.047	0.307	1.180	1.098	0.158
G35V									
avg distance (Å)	3.358	3.149	3.935(N^+) 4.677(N^0)	3.028(N^+) 4.365(N^0)	4.340	5.563	5.970	5.770	3.493

standard deviation	0.360	0.294	0.467 (N^+) 0.458(N^0)	0.407(N^+) 0.574 (N^0)	1.356	0.779	2.276	2.614	0.291
G35V/F71L									
avg distance (Å)	4.150	3.285	5.072(N^1) 3.833(N^0)	8.133(N^1) 7.444(N^0)	3.620	3.293	3.578	4.317	3.396
standard deviation	0.291	0.266	0.590(N^1) 0.404(N^0)	0.621(N^1) 0.614(N^0)	1.016	0.319	0.856	0.586	0.212
F71L									
avg distance (Å)	4.691	3.814	6.275(N^1) 6.858(N^0)	3.544(N^1) 5.606(N^0)	3.544	3.149	3.026	3.346	4.784
standard deviation	1.019	0.744	1.298(N^1) 1.307(N^0)	0.741(N^1) 0.816(N^0)	0.741	0.364	0.201	0.766	0.627
F71Y									
avg distance (Å)	4.001	3.015	3.825(N^1) 4.599(N^0)	3.313(N^1) 4.940 (N^0)	2.864	3.224	3.666	3.956	3.244
standard deviation	1.049	0.765	0.492(N^1) 0.385(N^0)	0.4917(N^1) 0.633 (N^0)	0.191	0.339	0.863	1.007	0.166

Table 2 shows the standard deviation and averages for the H-bonds distances shown in Figure 13 as measured in all our simulations. Besides examining the H-bond network within the α-helix we also examined the hydrogen bond interactions that formed the α-helix and Arg96. In wtGFP, the T62 and Y66 carbonyl oxygens H-bond to Arg96, allowing the Gly67 nitrogen to attack and form the chromophore, see Figure 2 in introduction .[30]

There was a significant difference in the distance between Cys70's carbonyl oxygen and Lys85's amide in the G31A and 1EMB structures, possibly due to the side chain of Lys 85 being

in a different conformation (Fig 14, 15). Other measured distances changed slightly, but overall the distances were similar with each other. There seemed to be some distances that had a greater change, however, these are not significant since they do not contribute to the chromophore maturation as they were not involved in the autocatalytic chromophore formation step or are part of the α-helix.

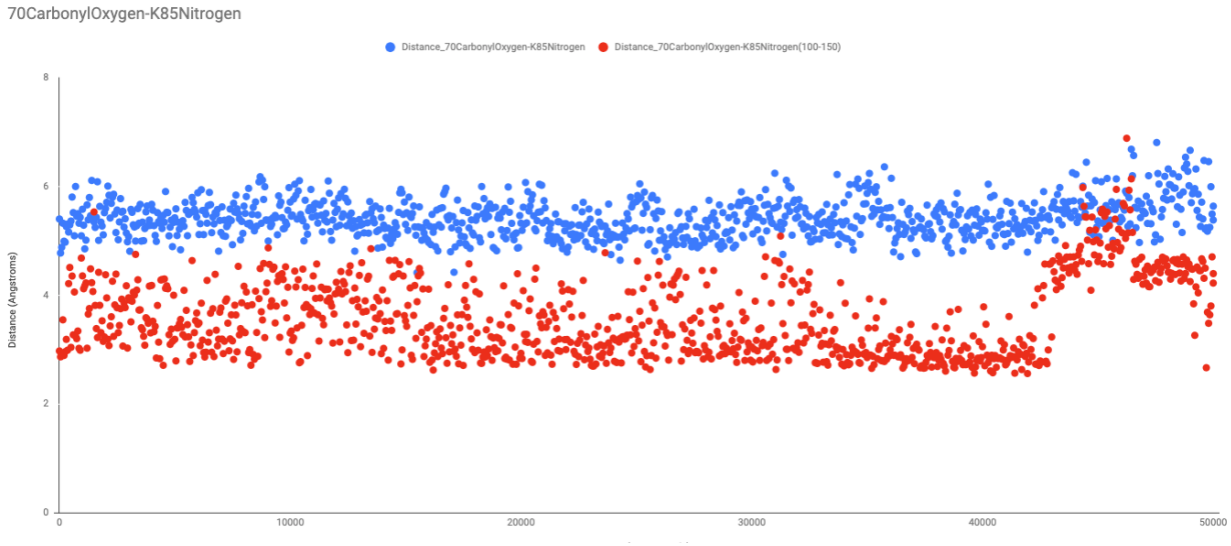


Figure 14. Distance between residue 70's carbonyl oxygen and K85's nitrogen.

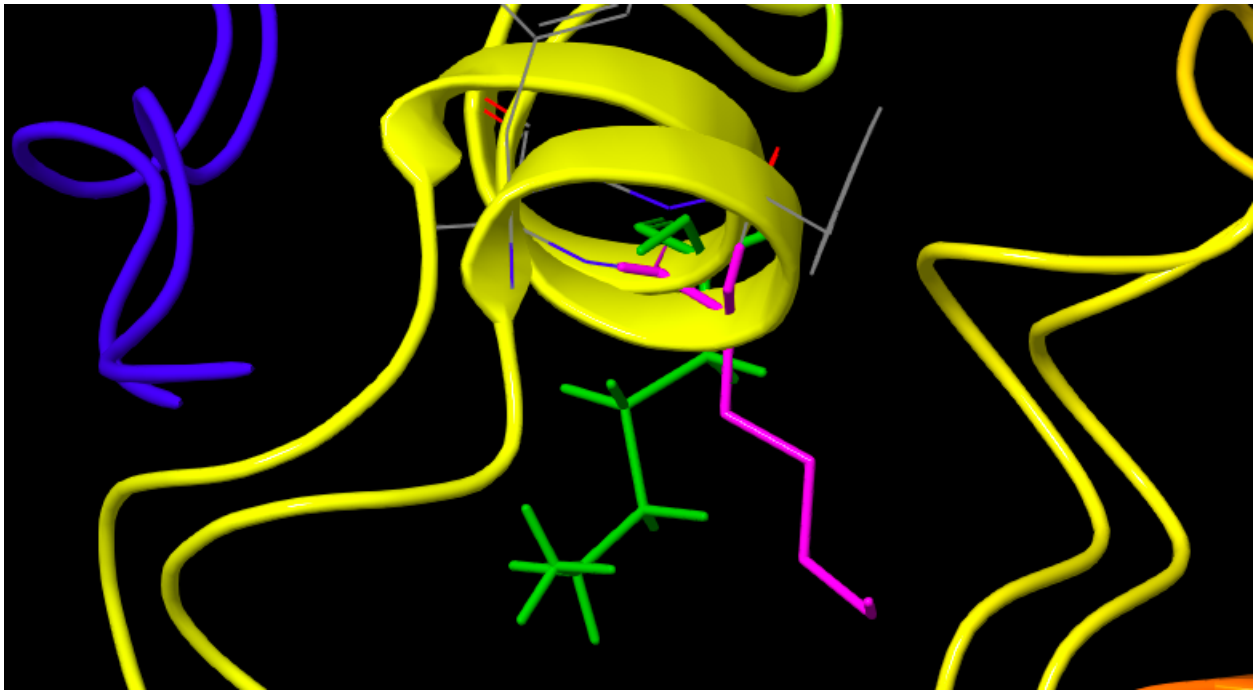


Fig 15. Visualization of the different conformations that Lys85 had in 1EMB (fuschia) and G31A (green), which accounted for the large difference.

The distance between the Arg96 and Tyr66 was vastly different amongst the glycine mutants. This distance increased by an average of 6 Å, potentially decreasing the catalytic effect that Arg96 has, slowing the chromophore maturation process. The increase could be caused by a change in a single dihedral angle and influence the formation of the chromophore. However, the exact cause is still unknown.

Interactions between aromatic rings and the conserved glycines of the second beta sheet:

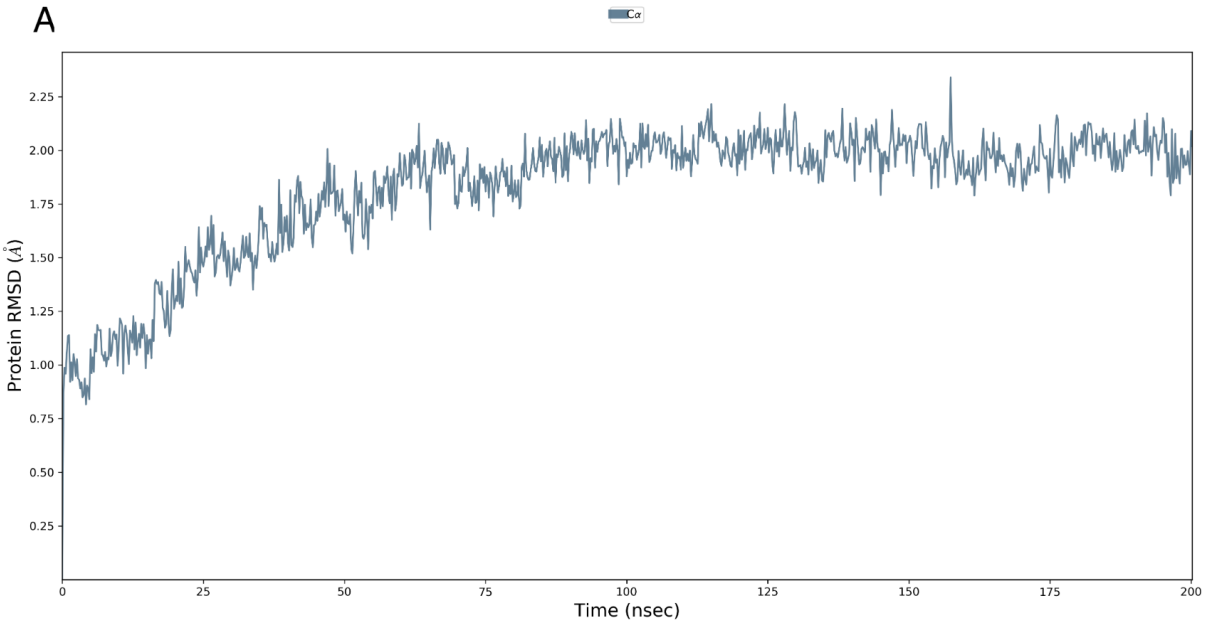
Earlier studies have noted that glycines present in parallel β-sheets are accompanied by phenylalanine across them on neighbouring β-sheets which provide a stabilizing effect.[31] This has also been observed in naturally occurring GFP structures. In wtGFP, two of the three glycines (Gly31 & Gly35) within the second β-strand have aromatic residues (Phe46 & Phe71, respectively) accompanying them. In order to investigate the effects that this interaction has on

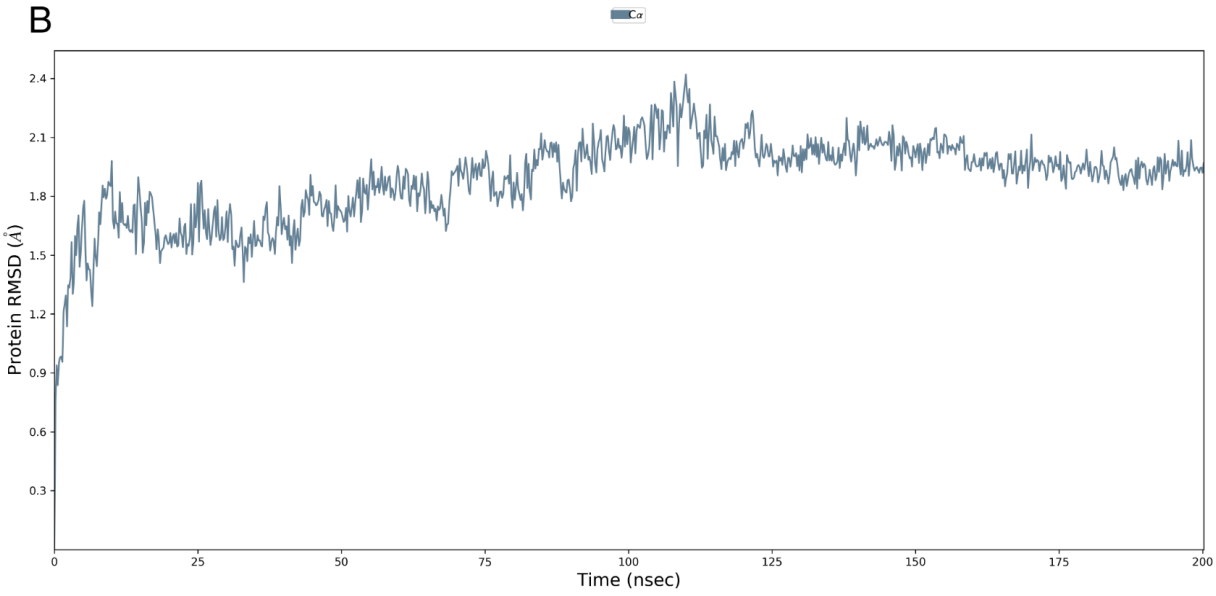
the spacing and overall structural stability, another batch of GFP mutants was engineered. These were mostly focused on the Gly35- Phe71 interaction, either increasing or decreasing the size of either residues as shown in Table 3.

Table 3. Mutations that were performed on the baseline structure, the engineered 2AWJ structure, and reasoning behind them.

Mutation	Explanation
G35/F71L	Decreased the size at position 71 in order to allow space within the structure.
G35V/F71	Investigate the increased steric effects and its effect on the G35 position.
G35/F71Y	Introduced a charge character in order to explore the effect on both the steric effects and the hydrogen bond network
G35V/F71L	Investigated a “balancing act,” where G35V mutation increased size and F71L mutation decreased it.

Structural validation was obtained in the same manner as described for the previous simulations, i.e. through the acquisition of a low RMSD. All of the simulations were able to reach an equilibrium within the completion of the 50ns simulation. Equilibration time was different for the four mutations, F71L, G35V, G35/F71Y, G35V/F71L, which were 38, 12, 5, 5ns, respectively (Fig. 16).


A
Cα
Protein RMSD (Å)
Time (nsec)
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
0
25
50
75
100
125
150
175
200


B
Cα
Protein RMSD (Å)
Time (nsec)
0.3
0.6
0.9
1.2
1.5
1.8
2.1
2.4
0
25
50
75
100
125
150
175
200

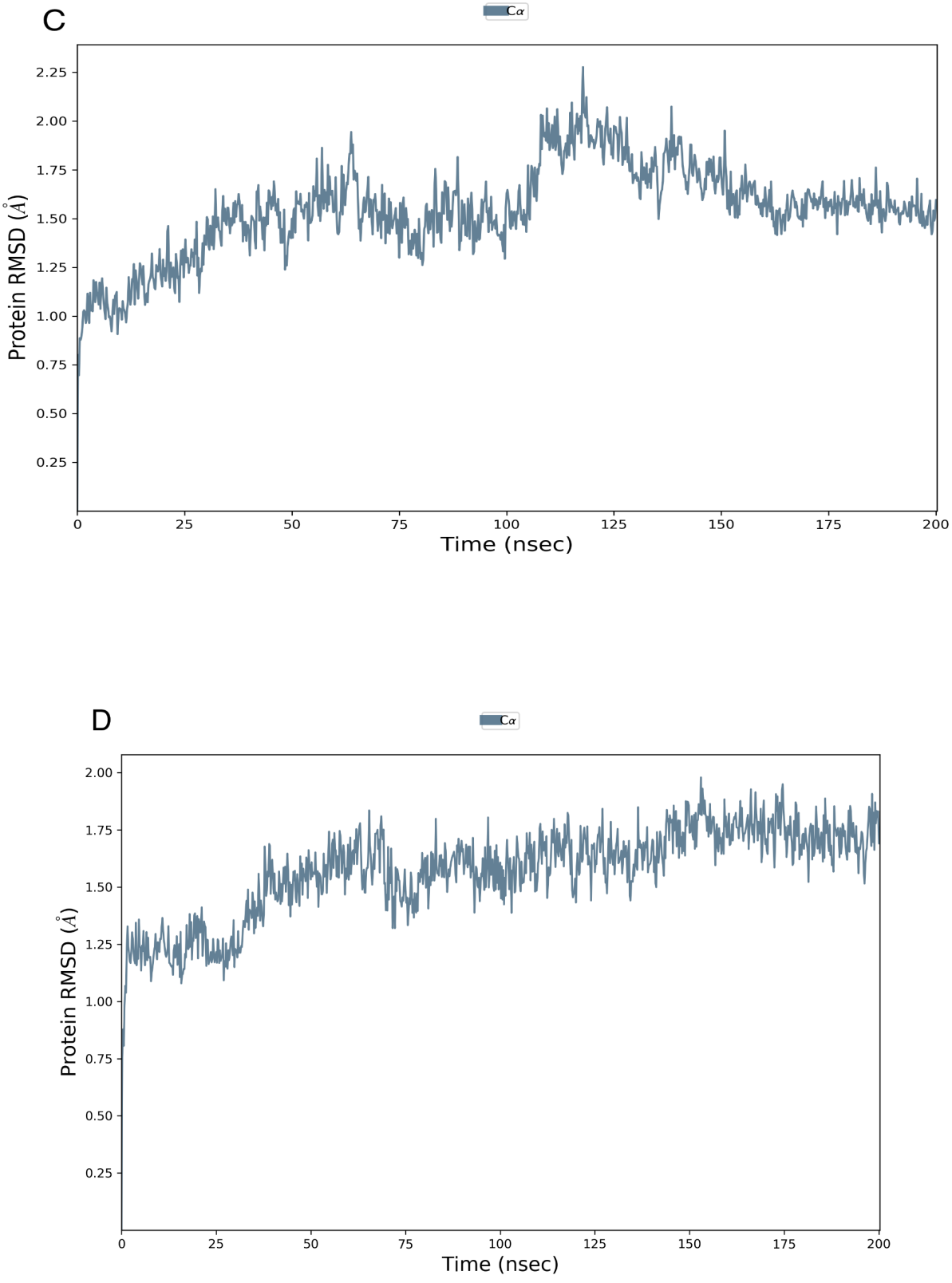


Fig 16. Graphical representation of RMSD of the (A) G35/F71L, (B) G35V/F71, (C) G35/F71Y, (D) G35V/F71L mutants.

Overall, the aromatic rescue interaction was expected for the non-mutated structure, thus a disruption or weakening of it was expected for the mutants. Structure integrity was expected to change with the mutants due to the steric clashes preventing it from reaching the normal conformation, and more than likely decreasing the rigidity of the barrel. The following measurements of the hydrogen bond distances between the backbone of the strands were taken to determine the impact of the glycine mutations, in which a bigger distance increased the rigidity of the overall structure (Table 4).

Table 4. Backbone hydrogen bond distances between β-strands 1-3, all in Å, for the hydrophobic pocket mutants.

Mutants	G35/F71L	G35/F71L	G35/F71Y	G35/F71Y	G35V/F71	G35V/F71	G35V/F71L	G35V/F71L	wtGFP	wtGFP
Dist,	Avg	St. Dev.	Avg	St. Dev.	Avg	St. Dev.	Avg	St. Dev.	Avg	St. Dev.
H25NH-V22CO	2.3708	0.4711	4.2686	1.8843	2.1462	0.2240	2.3196	0.3408	2.1590	0.2916
H25CO-V22NH	2.0995	0.2196	2.1546	0.2493	2.3045	0.3373	2.1432	0.2212	2.1590	0.2183
F27NH-G20CO	2.0567	0.2031	2.0743	0.2145	2.1061	0.2199	2.0957	0.2317	2.3746	0.2731
F27CO-G20NH	2.2558	0.2447	2.2935	0.2636	2.2911	0.2397	2.3205	0.2829	2.1289	0.2130
V29NH-L18CO	1.9240	0.1804	1.8914	0.1650	1.9841	0.1926	1.9773	0.1983	1.9780	0.2016
V29CO-L18NH	2.2748	0.3307	2.1845	0.3003	2.1206	0.2711	2.1588	0.3112	2.0395	0.2201

G31NH -V16C O	2.2785	0.3919	2.5359	0.9936	2.0481	0.2419	5.2081	0.4742	2.1886	0.3191
G31CO - V16NH	2.1016	0.3347	2.4322	0.7980	1.9987	0.2029	3.5411	0.5474	2.1522	0.2916
I47CO - S30NH	1.9772	0.1596	2.0311	0.1730	2.0319	0.1636	2.0398	0.1707	2.0425	0.1596
I47NH - S30CO	1.8992	0.1554	1.9443	0.1854	1.9278	0.1521	1.9215	0.1533	1.8994	0.1401
L45CO -E32NH	2.2052	0.3466	2.1875	0.3071	2.3183	0.4321	2.0648	0.2583	2.3074	0.3244
L45NH -E32CO	2.0901	0.2832	2.0857	0.2658	2.3377	0.5862	3.0888	0.5518	2.1230	0.2384
E34NH -T43CO	2.1241	0.2313	2.4578	0.7993	2.1788	0.4875	7.5276	0.9802	2.0301	0.1790
E34CO -T43NH	1.9961	0.1829	2.6536	0.8737	2.2030	0.7329	5.5800	0.5803	2.0023	0.1790
G35NH-V12CO	2.1886	0.2217	3.1598	0.8277	N/A	N/A	N/A	N/A	2.1107	0.2113
G35CH-V12NH	1.9271	0.1838	4.0419	1.8843	N/A	N/A	N/A	N/A	1.9454	0.1898
V35NH-V12CO	N/A	N/A	N/A	N/A	1.9669	0.1862	2.0892	0.2286	N/A	N/A
V35CO-V12NH	N/A	N/A	N/A	N/A	2.0694	0.2018	1.8852	0.1582	N/A	N/A

It was observed that there was an overall increase in the distances between the β strands, which aligned with the expectations caused by the increased steric effects. The G35V/F71L had the biggest increase between the H-bond distances of β-strands 1-3, with a maximum distance of 7.5276 Å between the hydrogen bond of E34NH and T43CO. For the G35V/F71 mutants the biggest distance was between Lys45NH and Glu32CO (2.3377 Å). All other distances fluctuated, but an overall increase was observed. A similar pattern was observed for the F71Y

mutation with the largest distance being between His25NH and Val22CO (4.2686 Å). The G35/F71L mutant was different - a slight decrease is observed, the largest distance was between Gly31NH and Val16CO (2.2785 Å) and biggest decreased compared to the wtGFP was 0.1023 Å.

H-bond distances were also observed for the glycine/alanine mutants and compared to those in the wtGFP (Table 5). As expected there was an overall increase in the distances between the strands, with the biggest distance corresponding to the G31A mutation between Ala31CO and Val16NH (3.0753 Å). This was expected due to the increased steric effects caused by the alanine's methyl group. The largest H-bonds distances in the G33A and G35A, were between Gly31CO-Val16NH (2.6427 Å) and Lys45CO-Glu32NH (2.5832 Å), respectively. These increases were not surprising due to the proximity that each of the pairs have to the mutated region.

Table 5. Backbone hydrogen bond distances for G3XA mutants, all in Å.

Mutants	G31A (100 - 150ns)	G31A (100 - 150ns)	G31A (150 - 200ns)	G31A (150 - 200ns)	G33A	G33A	G35A	G35A	wtGFP	wtGFP
Dist,	Avg	St. Dev.	Avg	St. Dev.	Avg	St. Dev.	Avg	St. Dev.	Avg	St. Dev.
L41CO - D36NH	2.6792	0.6415	2.4109	0.5911	2.1002	0.2483	2.2321	0.3019	2.0512	0.1868
E34CO - T43NH	2.4111	0.6908	2.0165	0.2182	1.9685	0.1640	2.0538	0.1952	2.0023	0.1790
E34NH - T43CO	2.6463	0.8218	2.1583	0.2757	2.0497	0.1856	2.0587	0.2256	2.0301	0.1747
L45NH	2.1728	0.4868	2.1469	0.3194	2.3117	0.3632	2.2523	0.3140	2.1230	0.2384

-E32CO										
L45CO-E32NH	2.3259	0.4949	2.3996	0.4111	2.5832	0.5980	2.4307	0.4495	2.3074	0.3244
I47NH-S30CO	1.9489	0.1681	1.9343	0.1486	1.8778	0.1546	1.8828	0.1575	1.8994	0.1401
I14CO-S30NH	2.0664	0.1808	2.0256	0.1701	2.0061	0.1503	2.0079	0.1573	2.0425	0.1596
A31CO-V16NH	3.0753	0.9130	2.5835	0.6149	N/A	N/A	N/A	N/A	N/A	N/A
A31NH-V16CO	2.8252	0.6864	2.6654	0.5162	N/A	N/A	N/A	N/A	N/A	N/A
H25NH-V22CO	2.5303	0.6867	2.2230	0.2998	2.3450	0.3910	2.3459	0.3488	2.1522	0.2916
H25CO-V22NH	2.1590	0.2357	2.0487	0.1909	2.1852	0.2372	2.1895	0.2481	2.1590	0.2183
V29NH-L18CO	1.9173	0.1882	1.9281	0.1892	1.8879	0.1580	1.9444	0.1882	1.9780	0.2016
V29CO-L18NH	2.3562	0.4192	2.4975	0.4567	2.0155	0.2013	2.0215	0.2461	2.0395	0.2201
G31NH-V16CO	N/A	N/A	N/A	N/A	2.5747	0.6116	2.2521	0.3916	2.1886	0.3191
G31CO-V16NH	N/A	N/A	N/A	N/A	2.6427	0.7767	2.1283	0.3049	2.1522	0.2916
I14CO-A33NH	N/A	N/A	N/A	N/A	2.2903	0.5172	N/A	N/A	N/A	N/A

I14NH-A33CO	N/A	N/A	N/A	N/A	1.9426	0.1839	N/A	N/A	N/A	N/A
F27CO-G20NH	2.2968	0.2481	2.0170	0.1880	2.3748	0.2714	2.2989	0.2558	2.1289	0.2130
F27NH-G20CO	2.0615	0.2046	2.1699	0.2246	2.1260	0.2283	2.0915	0.2085	2.3746	0.2731
G35CO-V12NH	1.9378	0.1911	1.9279	0.1878	2.0896	0.3482	N/A	N/A	1.9454	0.1898
G35NH-V12CO	2.1830	0.2489	2.1712	0.2697	2.2805	0.3447	N/A	N/A	2.1107	0.2113
G35CO-V12NH	N/A	N/A	N/A	N/A	N/A	N/A	1.9204	0.1679	N/A	N/A
G35NH-V12CO	N/A	N/A	N/A	N/A	N/A	N/A	2.1514	0.1981	N/A	N/A

Measurements were taken of the distance between the centroid of the aromatic ring and the α-carbon of the glycine in question. The centroid could not be selected directly thus the average distances of the 6 carbon members of the ring were taken and averaged to make a makeshift “centroid” for these measurements (Fig. 17). The main interaction that was explored through these measurements was that between Gly35 and Phe71, the other glycines were not examined because Gly33 is not accompanied by an aromatic residue.

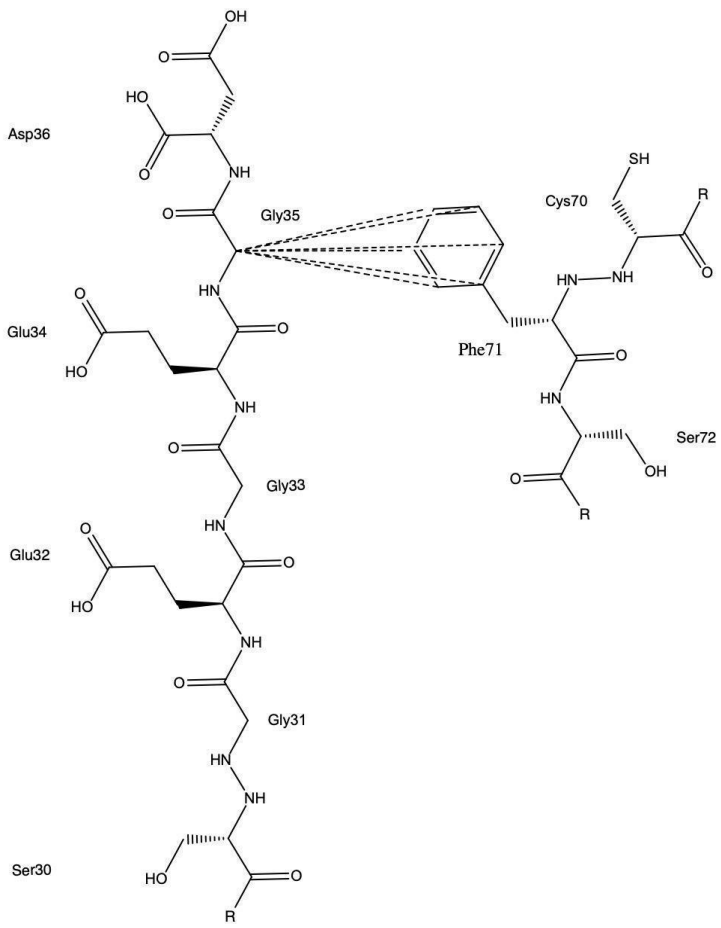


Figure 17. Visualization of the centroid measurements taken between Gly35 and Phe71 interaction.

These distances can affect the H-bond network across stranded pairs by either decreasing or completely eliminating their interactions.[31] There is a statistical preference in nature for glycines to be in the non-hydrogen bonding site and the phenylalanine in the hydrogen bonding site.[31] This preference results in the side chain of the phenylalanine residue to bend towards the α-carbon of the glycine, which could serve as a door allowing or blocking bulk solvent from entering the β-barrel.[31] The following distances between the targeted residues were obtained from the G3XA mutants and wtGFP structures (Table 6).

Table 6. Distance of F71 Phenyl Ring to G35 α-Carbon for G3XA Simulations

	Avg. F71- C_αG35 Distance (Å)	**Std. Deviation**
wtGFPimm	4.4796	0.4912
G31A (100-150)	5.0152	0.7044
G31A (150-200)	5.1193	0.7864
G33A	4.4332	0.4039
G35A	5.6834	0.4112

It was determined that the mutation increased the overall distances between the Phe71 and Gly35 with the average increase of 0.52 Å. Previous experiments suggest that the phenylalanine-glycine pairing provides a thermodynamic benefit for the structure, which aligns with the naturally occurring phenomenon.[31] The interaction with each other has possibly weakened as a result of the increased distance caused by the mutation, decreasing the observed thermodynamic benefit. Another possible effect caused by the increased distance is that it could have provided an entryway for waters into the GFP barrel since the residues would not have the same compactivity as the wtGFP. Either of these could have detrimental effects on structural stability and/or chromophore formation, possibly providing an insight on the functionality of the glycines.

Water Migration and Water Channels:

The effect that waters have on the maturation and cyclization of GFP species has been observed in previous studies and shown in the high conservation of waters in GFP's crystal strcuture.[15] Waters within the structure contribute to its H-bond network and thus can play an important role in the formation of the chromophore depending on their location and interactions (Fig 18). There are a number of mechanisms for chromophore formation that have been

proposed, but the following figure is the one that has been widely accepted (Figure 19). Waters are present throughout the process in order to stabilize the conformation needed for the cyclization.

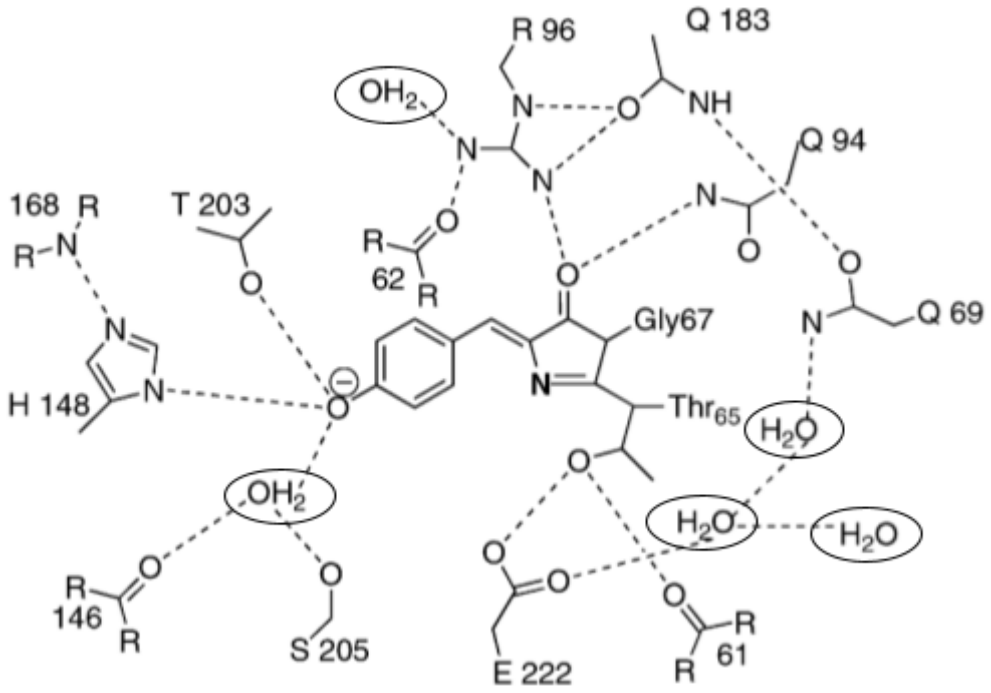


Figure 18. Visualization of the H-bonds that waters have around the chromophore and their interactions.[2]

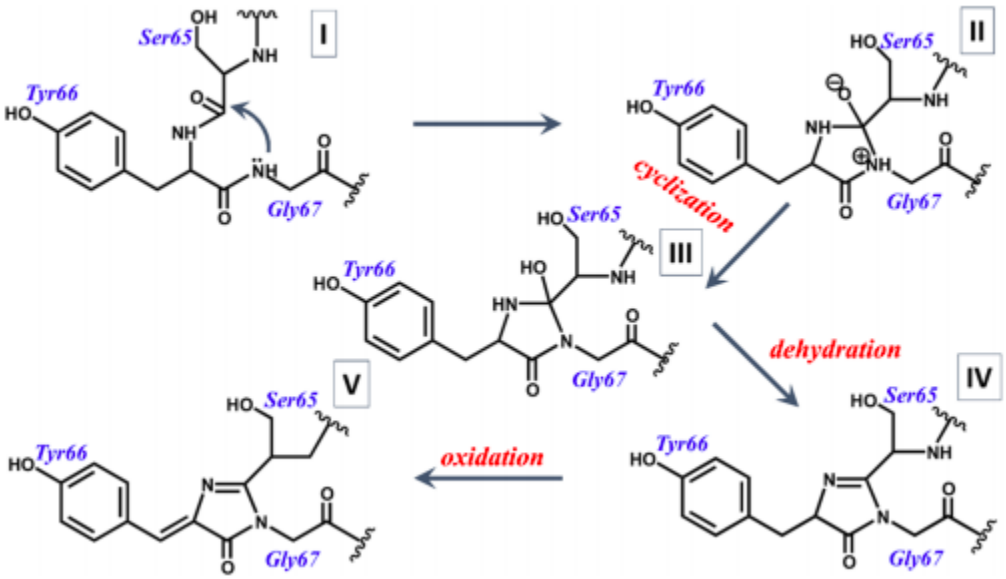


Figure 19. Chromophore forming mechanism that has been proposed, intermolecular auto cyclization interaction between the 3 residues: Ser65, Tyr66, and Gly67.[1]

There are waters within hydrogen bonding distance of residues Arg96 and Glu222, both of which are highly conserved catalytic residues within FPs.[32] In this aspect, the gap size between the β-strands of GFP can be significant to the overall cyclization of the chromophore since it will either allow or deny water access during chromophore formation.

Water migration channels were determined using the trajectory file and the MoleOnline program, a web-based application for determination of channels within structures.[32] Any water molecule that was within 5 Å of the structure was selected for trajectory analysis. In order to determine which waters traveled in and those that left the GFP barrel during the simulation, waters in the barrel in the first and the last frame were selected. These measurements were repeated for both simulations, 100-150 ns and the 150-200 ns.

Table 7. Waters that migrated either in or out of the protein in the G31A 100-150 ns with description of their movements.

Atom #	SPC #	Description	Entering Frame	Leaving frame
4842	404	Atom was on the edge (not completely in or out of the protein) at the beginning of the simulation in between the gap of ASN146 and the tenth beta strand, this water then moves in and stays within the protein for the rest of the simulation near the alpha helix THR62 (Yellow)	1	
23449	8273	This water is in between the 9th and 10th beta sheets, it left at frame 13 (in between the residues SER208 and MET218)		13
18201	4857	It is located in the middle of the lid, closest residue is LYS79, it left the protein at frame 31		31

4419	263	It is located near the previous water, however it is located closer to the residue SER72, it left the can at frame 42		42
4470	280	This water is in between the the first and the second beta strand near the residues GLY10 and ALA37, it left at frame 18		18
5958	776	The water entered into the protein in between residue PHE100 and ASN135 at the frame 984 where it is located for the rest of the simulation	984	
5307	559	The water is located in the middle of the lid of the GFP in between the residues LEU137, LYS131, and ASP103, it entered at frame 3 through the same passage that it resides in	3	
32634	9668	The water entered through the gap between the 10th and 11th beta strand, in between the residues HIE169 and TYR145, it is located in near the alpha helix residue THR62 at frame 664	664	
10944	2438	It entered through the same gap as the previous water at frame 969, this one is located near VAL150 at the end of the simulation, it made its way to the alpha and then moved near ASN149	969	
29901	8757	It entered through the lid gap that is located between GLU5 and GLN80 at frame 981 stayed relatively in the same area	981	
17775	4715	Entered the same gap as the previous at frame 995, it is near ALA37 at the end of the simulation	995	
13461	3277	Same gap as before at frame 999, and stayed within the same area	999	
27462	7944	Lid gap in between residues GLU5, LEU194, and ASP82 at frame 841, stayed in the same position for the rest of the simulation	841	
25653	7341	It entered the through the same gap (ALA37 and ARG73) at frame 818, stayed in this position for the 190 frames and then moved towards GLN80	818	
16884	4418	The final position of this water was near TYR75, entered through the gap of GLY4 and SER86 at frame 976, left and came back through GLU5, THR38 and LYS79 gap.(frame 995)	995	
15468	3946	Through the top lid at frame 999, stayed at the center	999	

18822	50664	Entered through the GLN80 and GLU5 gap (lid) and stayed there for the rest of the simulation		
27462	7944	Entered through the lid between LYS79 nad GLU5 at frame 840 stayed near ASP82 for the rest of the simulation		840
31326	9232	Entered through the lid in between ASP197 and ASP82 in the frame 999, stayed relatively in the same area		999

For the 100-150 ns simulation there were a total of 19 waters that entered into the G31A mutant. Of these 19 water molecules, 6 left the β-barrel at different times. There were a few waters that entered through the same gap between Ala37 and Gly10. This gap is not present in wtGFP, thus it is most likely the result of the mutation. Many waters entered through the lids of the residues. These waters usually stayed near the lids and did not seem to enter the barrel completely as a result of H-bonds formed on the termini. Waters that travel within the barrel remained near the α-helix throughout the simulation.

For the 150-200 ns simulation there were no waters that left the protein, all of them either entered or were already there. This analysis was repeated for the G33A (Table 9 & 10) and G35A (Table 11 & 12) mutants.

Table 8. These were the water that migrated either in or out of the protein in the G31A 150-200 ns with description of their movements. This is the continuation of the previous simulation thus water entered through the same gaps.

Atom #	SPC #	Description	Entering Frame	Leaving Frame
11728	2702	Starts close to the first couple residues of the protein (near GLY 10 and ALA 37). GLY 10 is on a turn while ALA 37 is on a 3/10 helix. Leaves from this space between frames 93 and 94 (around 4.60 ns into the simulation).		

28909	8429	Starts between LYS 131 and ASN 135 (on turn before 7th β-sheet). Definite H-bonding with the oxygen of the water and amino group off of the side chain of each residue with possible H-bonding with amide nitrogen of ASN 135 backbone. Stays in the same position for the whole simulation.		
28909	8429	Entered and stayed in the same position for the entire simulation, minimal movement around LEU137	5	
21277	5885	Entered and stayed in the same position for the rest of the simulation, it is positioned near LYS131	804	
7003	1127	Entered through the gap ASP102 and GLY134	986	
19771	5383	Entered through the gap LEU53 on the lid of the barrel, final position is in between VAL55 and HIS217	53	
8464	1614	Entered through the gap of ASN146 and SER205, final position near the same gap, minimal movement	470	
21265	5881	Entered through the gap of ASN146 and SER205, final position near the same gap, minimal movement	480	
18586	4988	Entered through the same gap as before (ASN146 and SER205), near the same gap, minimal movement	101	
29761	8713	Entered through the same gap as before (ASN146 and SER205), final position between TYR66 and LEU44	20	
17623	4667	Entered through a gap in the lid of the barrel (between ASP82, ASN198 and GLY228), made its way down the beta barrel ended near TYR66	355	
22585	6321	Entered through a gap formed by VAL163 and ILE152, final position near HIE81	578	
17176	4518	Entered through a gap formed by HIS181 and THR38 in the lid of the barrel, final position near CYS70 in the lid	679	
13981	3453	Entered through a gap formed by LYS85 and GLY4,	928	

		final position near ARG73 in the lid		
15667	4015	Entered through a gap formed by GLY4 and SER86, stayed in this position for the rest of simulation with minimal movement	10	
24841	7073	Entered through a gap formed by GLU5 and LYS85 and stayed with minimal position, final position near HIE81	960	

Table 9. Waters that were within the ß-barrel at frame 1 of the G33A simulation.

Atom #	SPC #	Description
57679	15027	The starting position for this water was between Y143 and H169, within the first 20 frames of the simulations, it flowed out. It never re-entered the barrel.
57238	17880	The starting position for this water was between Y143 and H169, within the first 20 frames of the simulations, it flowed out. It never re-entered the barrel.
57451	17951	The starting position for this water was between Y143 and H169, within the first 20 frames of the simulations, it flowed out. It never re-entered the barrel.
57682	18028	The starting position for this water was between Y143 and H169, within the first 20 frames of the simulations, it flowed out. It never re-entered the barrel.
57532	17978	Positioned near D149 and S205 in the first frames, moving towards the α-helix as the simulation progressed. It stayed near the α-helix until fram 295, where it left through the gap between N146 and A206.
57151	17851	Positioned near the α-helix and β-strands, between L60 and H181. Roams around this area until frame 55, where it leaves through the gap formed by Y145 and N170.
57445	17949	Positioned near the α-helix between T62, T59 and I167. Moves towards the V61 & L60 of the α-helix and stayed there for the rest of the simulation.
57625	18009	Positioned near the L201 and Y66, between the α-helix and β-strand. Moves closer to the α-helix near S65 and roams

		around the area between G67 and C70, normally staying closer to S65. It then slowly moves out of the protein through the gap between S147 and A206 in frame 181. It does not enter the protein after this.
57154	17852	Positioned between T62 and V61. The water stayed near the V61 of the α-helix, where it stayed for the rest of the simulation.
57277	17893	Positioned near the S65 of the α-helix. It moved towards the Q69 and C70. It was later pushed off and it moved towards the strands for a bit after coming closer again to the α-helix.
57559	17987	Positioned near G67 of the α-helix. It moves closer to the α-helix, roaming the area of the chromophore forming residues. It started to head out, moving towards the lids of the barrel.
57736	18046	Positioned near the T108 and E124. Water moves towards the Y66 and roams around the chromophore forming residues.
57531	17911	Starting near the R96, it moved a little towards the chromophore forming region towards the G67, and it stayed in that area for the rest of the simulation.
57160	17884	Positioned near the E5, L85, and C70. It stayed around this area for the entire simulation, moving back and forth between the surrounding residues.
57187	17863	Positioned near the F84, D197. It stayed around this area for the entire simulation, moving back and forth between the surrounding residues.

Table 10. Waters that were within the ß-barrel at frame 1001 of the G33A simulation.

Atom #	**SPC #**	**Description**
14062	3488	Ended near the opening of S205 and other waters, but the others did not completely enter the structure. Entered at frame 986 between the gap S197 and S205. Where it moved towards the α-helix.
26608	7670	Ended near α-helix close to T62. Entered through the gap formed by S147 and S205 at frame 977.
12805	3069	Ended near the α-helix by P58. Entered in the 266 frame between the gap friend by S205 and Y143, and it moved

		towards the L60, where it stayed near it for the rest of the simulation.
22537	6513	Ended near the lid residues by N144. Entered at frame 226 between the gap N144 and H169. It moved towards the general direction of E142, where it stayed at.

Table 11. Waters that were within the ß-barrel at frame 1 of the G35A simulation, courtesy of Justin Nwafor.

Atom #	SPC #	Description
55639	17318	• Begins nearest to Phe 8 on α-helix position of strand 1 • Leaves at frame 4 from termini capped end of barrel
55195	17170	• Begins closest to Phe 84 on the α-helix that runs through the center of barrel • Leaves at frame 403 (40.3ns) from the termini capped end of barrel
55222	17179	• Begins closest to Phe 83 on the α-helix that runs through the center of barrel • Leaves at frame 695 (69.5ns) from the termini capped end of barrel
55312[ff]	17209	• Begins closest to Phe 71 near α-helix that runs down the center of the barrel • Moves toward Gly 67 at frame 142 (14.2ns), but never leaves cavity during the simulation
55660[ff]	17325	• Begins closest to Phe 71 near α-helix that runs down the center of the barrel • Moves toward Gly 67 at frame 142 (14.2ns), but never leaves cavity during the simulation
55771[ff]	17362	• Begins closest to Gly 67 near α-helix that runs down the center of the barrel • Remains there over the course of the simulation, never leaving the barrel
55186[ff]	17167	• Begins closest Leu 60 near α-helix that runs down the center of the barrel • Moves closer to Ala 179 at frame 792 (79.2ns) and stays there inside barrel for the remainder of the simulation
55567	17294	• Begins closest to Ser 205 on Strand 9 • Leaves at frame 467 (46.7ns) through strands 9 and 10

55189	17168	• Begins closest Thr 62 near α-helix • Leaves through strands 7 and 8 at frame 8
55486	17267	• Begins closest to His 169 near strands 7 and 8 • Leaves through strands 7 and 8 at frame 8
55234[ǂ]	17183	• Begins closest to Asn 135 on α-helix on the end of the barrel without the termini • Remains there over the course of the simulation, never leaving the barrel
55273	17196	• Begins closest to Tyr 145 between strands 6 and 7 • Leaves through strands 6 and 7 at frame 2
55594[ǂ]	17303	• Begins simulation closest to Gln 69 on α-helix that runs down the center of the barrel • Remains there over the course of the simulation, never leaving the barrel
55246	17187	• Begins closest to Asn 170 on strand 6 on the end of the barrel without termini • Leaves at frame 229 (22.9ns) between strands 7 and 8

Table 12. Waters that were within the ß-barrel at frame 1001 of the G35A simulation, courtesy of Justin Nwafor.

Atom #	**SPC #**	**Description**
53521	16612	• At frame 1001, this water molecule is closest to K101 on the end of the barrel without termini • Water molecule enters GFP at frame 882 through the loops next to K101
22708	6341	• At final frame, water molecule is closest to Ile 171 between strands 6 and 7 • Enters barrel at frame 767 through the end of the barrel without termini
16060	4125	• At frame 1001, water molecule is closest to S147 between strands 5 and 6 • Enters cavity at frame 928 from end of GFP without termini through strands 5 and 6

37009	11108	• At final frame 1001, water molecule is closest to F84 on α-helix that runs through the center of barrel on the end without the termini • Enters cavity at frame 335 next to the on α-helix that runs through the center of barrel on the end with termini closest to R73
38710	11675	• At last frame, water molecule is closest to Y66 on the on α-helix that runs through the center of barrel, right in the center • Enters cavity at frame 237 through a gap between strands 6 and 7
19321	5212	• At final frame, water molecule is closest to D36 on the second strand • Enters GFP at frame 951 through the top of GFP with termini
32431	9582	• At final frame, water molecule is closest to Lys 85 on α-helix that runs through the center of barrel on the end with the termini • Enters barrel at frame 874 through the termini capped end of GFP
23893	6736	• At frame 1001, water molecule is closest to Ile 188 at the termini capped end of GFP • Enters GFP at frame 888 through termini capped end

[ǂ] - Stays in the same position for the whole/rest of simulation.

Due to minimal water movement observed within the β-barrel, it is suspected that these waters hydrogen bonded to something within the β-barrel, whether a residue within the α-helix or the β-barrel itself. This hypothesis explains the minimal movement that is observed in the simulations and the area that the waters inhabited (near the α-helix). As well as the fact that crystal structures of fluorescent proteins have numerous well defined waters, some of which are conserved across many structures. There were waters that moved a significant amount towards the α-helix and remained there until the end of the simulation. The increased water migration is observed to be focused around the “lid” area of the barrel, which is expected as it is exposed to the bulk solvent and their secondary structures being composed of loops. It is observed that there is an overall decrease of water migration on the area around the α-helix which could be the result of the residues increasing distance between the strands, increasing the rigidity of the barrel. The

increase in steric effects caused by the alanine's methyl could be disrupting the H-bond network within the GFP, distorting the spaces between the strands, thus making it harder for waters, see *Interactions between aromatic rings and the conserved glycines of the second beta sheet,* pg 31.

Water channels were calculated through MOLEonline to draw the comparison between the observed water movement and predicted channels. From the possible channels determined by the program, a structural comparison can be drawn between the predicted channels and the observed water movement, as well as correlations between the dimensions of the water channels (Fig 20 & 21, and Tables 12-16). It was observed that water channels were concentrated in the lids area of the β-barrel, possibly due to the increased steric effects of the alanine causing smaller spacings between the β strands. The lack of waters observed within the barrel of the mutants aligns to the observed decrease of channels.

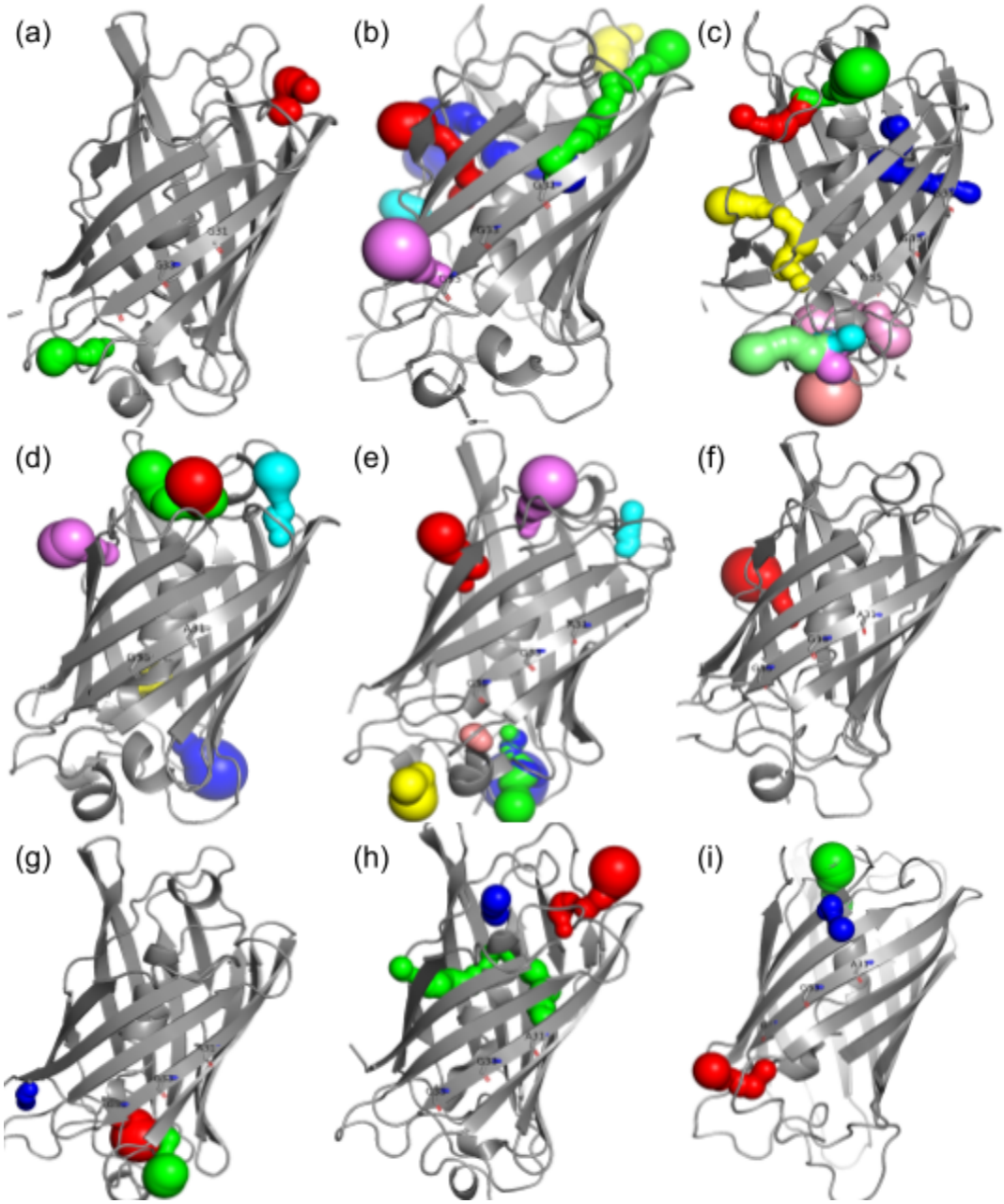


Figure 20. Visual representation on the predicted water channels calculated through MoleOnline in the immature wtGFP (first row, a-c) and G31A mutant, middle (d-f, 100-150ns simulation) and last row (g-i, 150ns-200ns simulation).

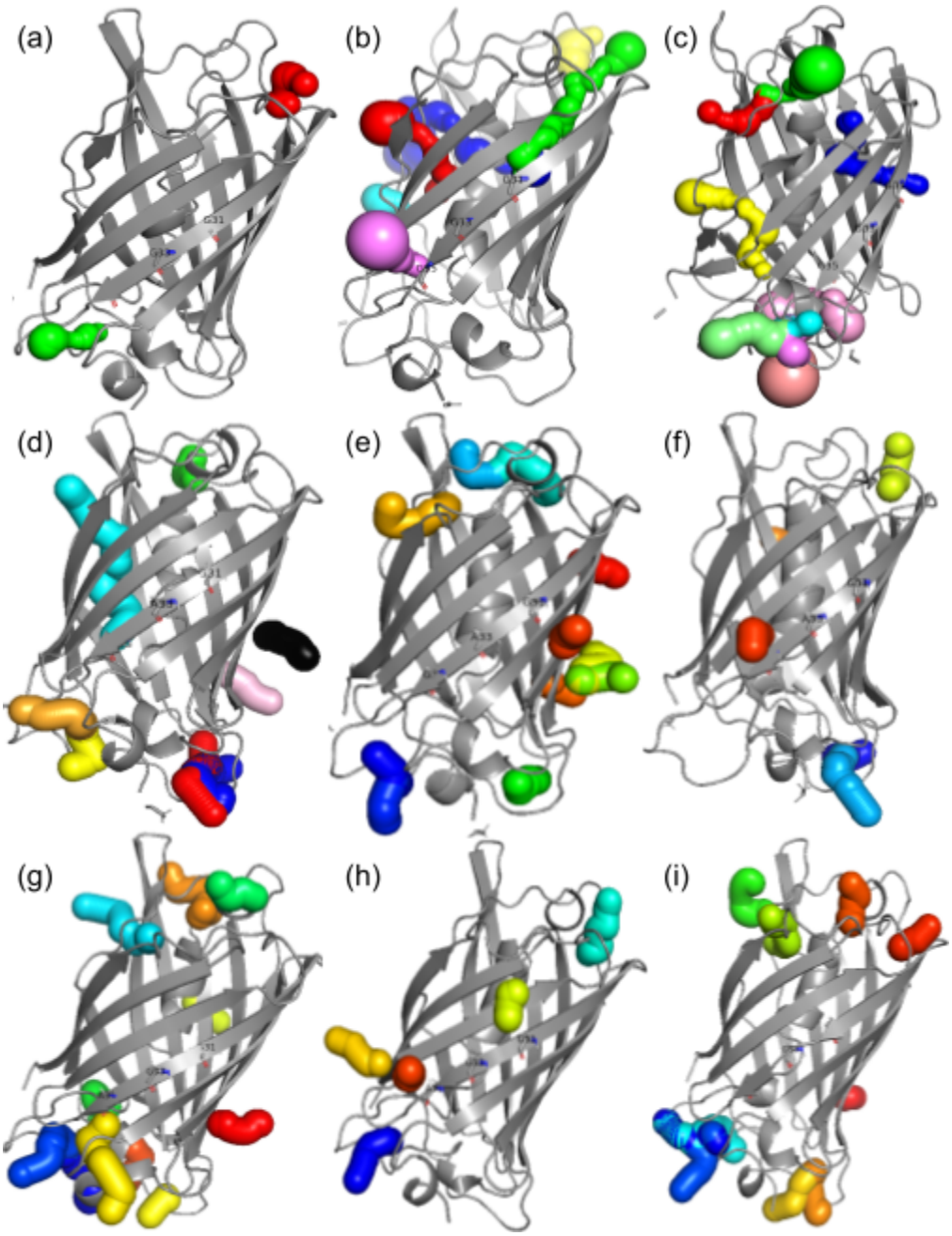


Figure 21. Visual representation on the predicted water channels calculated through MoleOnline in the immature wtGFP (first row, a-c) and G33A mutant, middle (d-f) and last row (g-i).

Table 12. MOLEonline's predicted channels and their dimensions for the wtGFP simulation, courtesy of Justin Nwafor

Frame	**Name**	**Length (Å)**	**Bottleneck (Å)**	**Lining Residues**	**Description**
1*	T1C3	8.2	1.5	V22, H25, P54,	Located at the end opposite of

				L137, ***V22***	the C and N termini. Near the 1st beta sheet and the turn after the 6th beta sheet. It has something of an U-shape.
1*	T2C4	8.2	1	D36, A37, T38, Y39, ***F8, T9, G10, A37***	Located near the C and N termini, near the 1st and second beta sheets.
501	T1C2	17	1.1	V61, T62, S65, Y66, N144, Y145, T203, S205, L207, L220, ***V61***, ***T62***, ***Y145***, ***S205***, ***A206***, ***L207***	Located near the alpha helix, exiting between the turn of the 7th beta sheet and the 10th beta sheet.
501	T2C2	29	1	L18, V22, N23, H25, F27, V29, L53, P54, V55, T57, L60, T63, F64, I123, L125, E132, L137, ***V22***, ***N23***	Located near alpha helix, exits between the turn of the 1st beta strand and the turn after the 6th beta sheet.
501	T3C2	39.5	0.8	L60, V61, T62, T63, F64, R96, I98, Y106, I123, L125, Y145, N146, H148, R168, H169, N170,H181, ***R168***, ***H169***, ***P58***, ***T59***, ***L60*** ***N144***, ***Y145***, ***N146***, ***K166***	Located near alpha helix, exits near the 7th beta sheet. It looks as if it is connected to the tunnel, T2C2. However, they're oriented in opposite directions.
501	T4C3	7.7	2	D129, K131, D133, ***D102***, ***D103***, ***D129***, ***G134***	This tunnel is almost completely outside of the beta barrel. It is located opposite of the C and Termini end of the protein near the 4th and 6th beta strands. It's also very compact.
501	T5C4↥	9.4	1.1	D36, L42, T43, F71, ***E34***, ***G35***, ***D36***, ***K41***, ***L42***	Near the 2nd and 3rd beta sheet and F71 of the alpha helix. G35's backbone also lines this cavity.

501	T6C5	3.8	1.3	S147, N149, T203, ***H148***, ***L201***, ***S202***, ***T203***	This tunnel is only lined by beta sheet residues, it does not go far enough into the barrel. Exits between the 10th and 7th beta strands.
1001	T1C1	11	1.1	P58, L141, Y145, N146, H169, N170, ***E142***, ***N144***, ***Y145***, ***N146***, ***R168***, ***H169***, ***N170***	Only lined by alpha helix residue (P58). Leaves between the turn of the 6th beta sheet and the residues before the 5th beta sheet.
1001	T2C1	12.7	1	P58, H139, L141, Y143, H169, K209, ***L141, E142 , N170***	Seems to be connected to tunnel T1C1. It is lined by some of the same residues as T1C1, but it moves out more in the opposite direction, allowing it to be lined with residues that are near the turn of the 10th and 11th residue (i.e. K209).
1001	T3C1	26.9	1.1	V16, E17, L18, F27, V29, F46, L53, V55, L60, T63, R96 , T108, I123, L125, ***V16 E17***, ***L18***, ***V29***, ***S30***, ***E124***	This tunnel is wide enough and twists within the beta barrel, resulting in it being near residues that are on 6 different beta sheets, and alpha helix. Exits between the 1st and 2nd beta sheets, slightly above G31.
1001	T4C2	24.3	1	Y66, Q69, S72, Y74, F84, V150, Y151, I152, I161, V163, F165, N185, H199, L201, ***Y66***	Closer to the C and N termini of the barrel, very close to the alpha helix. Exits through 7th and 8th beta sheets.
1001	T5C3	9.9	1.2	E5, F8, T9, A37, T38, K85, ***G4, F8, T9, A37, T38***	Smaller tunnel, almost completely outside of the barrel. It's located almost right next to the C Termini of the protein.
1001	T6C3	10.7	1.2	E5, F8, A37, T38, K85, ***G4, F8, T9, A37, T38***	There's almost complete overlap with T5C3. Their directions differe, in that they point at the outermost point of the tunnels. This tunnel points back up between the two helices while

					T5C3 points to the side near the C termini.
1001	T7C3	14.4	1.3	K3, E5, F8, K85, S86, L194, ***G4, E5, K79, Q80, D82, S86***	This tunnel looks connected to both T6C3 and T5C3, but it points in a separate direction. This one actually does pass by the C termini while the other two just approach it. This tunnel is also much wider than the two that were previously described.
1001	T8C3	15.6	1.2	E5, F8, A37, T38, R73, K79, K85, ***G4, F8, A37, T38, Y74, D76***	This tunnel also looks connected to the three previously described tunnels, but this one points opposite the direction of T5C3 for a longer length which allows it to be lined with residues like Y74.
1001	T9C4↟	21.5	1.2	E90, K156, N159, P187, P192, V193, L195, ***S86, A87, G189, D190, G191, P192, V193***	This tunnel is located on the bend after the 9th beta sheet. The middle of it sits on the helix and then each side the tunnel goes out of the protein. The side closest to the C termini gets in fairly close proximity to T7C3.

Table 13. MOLEonline's predicted channels and their dimensions for the 2AWJ G31A simulation (100-150ns section), courtesy of Justin Nwafor

Frame	**Name**	**Length (Å)**	**Bottleneck (Å)**	**Lining Residues**	**Description**
1	T6C9	8.9	1.1	V22, H25, F27, P54, V55, L137, ***V22, H25, P54***	This tunnel is located on the end of the barrel opposite of the C and N termini. It is lined by residues of the end and turn of the first β-strand, the turn between the sixth and seventh β-strands, and residues that are close in sequence to the alpha helix in the barrel.
1	T3C3	6.5	1.6	F83, A154, P187,V193, L195, **K156,**	This tunnel is located on the lid of the beta barrel on the side of K158

				K158, G160	and G191. It is on the turns of the seventh and eighth beta sheet.
1	T4C4	16	1	Y66, Q69, F84, I152, M153, I161, V163, Q183, N185, L201, **Y151, I152, M153, K162**	The tunnel travels through the residue in between I152 and V163. It is in between the 6th and 7th beta sheet. It does not get sufficiently close to the alpha carbon.
1	T5C6	6.8	1.2	N144, Q204, A206, L207, **Y145, S205, A206, L207**	The tunnel slightly enters the beta barrel through the gap between L207 and Y145. It is a gap between the turn of the 6th and 7th beta sheet, and the 10th beta sheet.
1	T2C2	11.3	1.6	K52, P56, W57, P58, H139, Y143, E172, D216, ***L53, V55, P56, W57, L141, E142***	The tunnel is in one of the lids of the beta barrel on the side of E142. It seems to enter the protein slightly before leaving yet again. It overlaps with the tunnel T1C2
1	T1C2	3.4	2.2	K52, W57, H139, D216, ***L53, V55***	The closest residue is L53, and it overlaps with T2C2. It seems not to enter the protein.
501	T3C2	20.7	1	F83, A87, Y92, N159, P187, V193, ***A87, G91, K156, P187, I188, G189, D190, G191***	The tunnel goes through the V193 and I188. The bottleneck is bending towards F84.
501	T2C2	13.2	1.1	F83, A87, E90, Y92, P187, V193, ***S86, A87, G91, P187, I188, G191, P192***	It is perpendicular to the tunnel T3C2 and travels in the same manner as it.
501	T4C3	8.9	1.6	K3, E5, K79, L194	This tunnel is located in the center of the lid that contains G4. It does not enter the protein, it remains entirely outside.

501	T1C1	11	1.1	V61, N144, Y145, S205, L207, L220, ***Y145, N146, S205, L207***	This tunnel enters the protein through a gap provided by Y145 and A206. It bends towards the alpha helix slightly, but does not get too close to it.
501	T5C4	10.2	1.1	P56, P58, H139, Y143, H169, K209, ***P56, T59, L141, E142, Y143***	The widest part of the tunnel is outside the protein on the lid that contains E142. It enters in between W57 and E142. P58 is the residue that is closest to the end of the tunnel.
1001	T1C3	10.9	1.1	Y66, H148, N149, V150, F165, R168, ***S147, H148, N149, K166***	The tunnel is located between residues V150 and F165.

Table 14. MOLEonline's predicted channels and their dimensions for the 2AWJ G31A simulation (150-200ns section), courtesy of Justin Nwafor

Frame	**Name**	**Length (Å)**	**Bottleneck (Å)**	**Lining Residues**	**Description**
1	T2C6	9.1	1	V12, P13, F114, D117, L119, **L7, V11, P13, D117, L119**	The tunnel is located between the D117 and V12, in between the second beta sheet and the turn of the 5th beta sheet.
1	T3C7	4	1.1	M78, H81, H199, I229, **N198, G228**	The tunnel is located between G228, N198 and H199. It is between the 10th and the 11th beta sheet
1	T1C5	8.1	1.5	F83, N159, P187, V193, P196, **K158, V193, L194**	The tunnel is located between L194, T186, which means that it is close to the 9th beta strand and the turn that connects both the 9th and the 10th beta strand.
501	T1C1	15.1	1	V22, H25, P54, V55, Y106, F130, E132, L137**, V22, N23, P54**	This tunnel is located in between the barrel lids in between N23 and L53. The

					bottleneck resides within the protein, but as the tunnel moves outward it widens.
501	T3C4	5.5	1.2	P56, P58, T59, Y143, H169, T59, **L141, E142, Y143**	The tunnel is located between W57 and E142, outside the beta sheets, near the lids. This tunnel is in the same lid as the T1C1 of this frame.
501	T2C1	44.1	0.9	V16, L18, V29, F46, L60, T62, T63, F64, S65, Y66, I98, F100, Y106, Y108, I123, L125, Y145, S147, H181, T203, Q204, S205, E222, **S30, T59, T62, E124, N146, S147, T203, Q204, S205**	The entry/exit is in between the residues Q204 and N146, in between the 10th and 7th beta sheet. It loops around the alpha helix on the side of V162, Y182, I98 , L125, L18, ending in S30. It seems to have equal distance between the alpha helix and the beta sheet.
1001	T2C3	9.6	1.1	P56, P58, T59, H139, L141, Y143, H169, **W57, T59, L141, E142**	The tunnel is located L141 and P58. It is located in the lid and the entry to the protein is located near the alpha helix.
1001	T3C4	9.7	0.9	E32, K45, I47, R215, H217, **F46, M218**	The tunnel is outside the protein and it does not enter anywhere. The closest the tunnel is in the protein is at F46 and M218.
1001	T1C1	11.5	1.2	V11, V12, D36, A37, T38, **E6, L7, F8, T9, G10, D36, A37, T38**	It is located in the other lid, opposite to the first tunnel described for this frame. It is in between G10 and A37. It goes slightly in towards the bottleneck

Table 15. MOLEonline's predicted channels and their dimensions for the 2AWJ G33A simulation, courtesy of Justin Nwafor**.**

Frame	**Name**	**Length**	**Bottleneck**	**Lining Residues**	**Description**

		(Å)	(Å)		
1	T1C1	28.3	0.9	L42, V61, Y66, Q69, S72, N144, Y145, T203, S205, A206, L207, L220, E22, V224, **S65, Y145, L207**	Comes in between strands 10 and 7 opposite the N/C termini, then goes all the way toward the chromophore tripeptide and other parts of the α-helix.
1	T2C3	17.1	1.3	A87, E90, N159, P187, G189, D190, G191, V193, **S86, G189, D190, G191**	Mainly lined by loop regions between β(9-10) and β(7-8) on the same side as the N/C termini. Somewhat horseshoe shaped.
1	T3C3	18.8	1.2	A87, P89, E90, N159, P187, G189, D190, G191, V193, **S86, G189, G191**	Overlaps significantly with T2C3. Close to the N-terminus.
1	T4C5	7.1	1.2	K3, E5, K79, D82, L194, **K79**	Small tunnel lined by residues of the N-terminus, loop between β(9-10) and the end of the α-helix.
1	T5C5	13.1	1.1	E5, A37, T38, R73, Y74, P75, K79, D82, K85, **A37**	Slight overlap with T4C5, also located close to N-terminus, closer to the end of the α-helix, and interacts with residues of the β-turn between strands 2 and 3.
1	T6C6	6.5	1.3	R109, A110, E111, R122, I123, E124, **E111, R122, E124**	Lined by residues pointing outward toward bulk solvent. Ends at space between strands 5 and 6, almost aligned with the chromophore tripeptide, but does not go into the barrel to interact with those residues.
1	T7C7	5.8	1.1	D102, D103, K131, G134, N135, I136, N177, **K101, D102, N135**	Lined by residues of loops between strands 4,5, and the lid opposite the N/C termini.
1	T8C8	6.1	1.3	V93, E95, K158, T186	Tunnel points toward the space between β-strands 4 and 9, but does not actually enter the β-barrel.

501	T1C2	7.1	1.8	E5, T9, A37, T38, K79, Q80, **E5, F8**	Somewhat bean-shaped, lined by residues close to the N-terminus, the β-turn between strands 2 and 3, and residues in the loop between the α-helix and strand 4.
501	T2C2	8.4	1.8	E5, E6, T9, K79, Q80	Some overlap with T1C2, points further out into the bulk solvent.
501	T3C3	7.4	1.8	D102, N135, I171, S175, Q177, **K101, V176**	Small tunnel lined by residues on loops between strands 4 and 5 and the lid opposite the N/C termini.
501	T4C3	13.8	1.1	D102, D103, F130, K131, K131, G134, N135, Q177, **D102, D129**	Some overlap with T3C3, points in the opposite direction when exiting the β-barrel.
501	T5C6	8.2	1	E90, P187, G189, D190, G191, V193, **S86, G189, D190, G191**	Lined by residues on the loop between strands 9/10 and the loop between the α-helix and strand 4.
501	T6C7	6.4	1.4	E111, K113, V120, R122, **E111**	Points toward the space between strands 5 and 6, but does not enter the β-barrel.
501	T7C7	11.1	1.4	R109, A110, E111, K113, V120, R122, E124, **E111, R122**	Overlap with T6C7, goes in the opposite direction on the way out of the β-barrel.
501	T8C9	11.9	0.8	P58, Y143, N144, Y145, H169, L207, **E142, Y143, N144, Y145**	Lined by residues at the top of the α-helix (opposite N/C termini). Exits the β-barrel between strands 7 and 10.
501	T9C10	6.6	1.2	V93, E95, Q184, N185, T186, **E95, Q184**	V-shaped, points toward the space between strands 4 and 9, but never enters the barrel.
501	T10C11	4	1.1	L15, E17, S30, R122, **V16**	Points between strands 1 and 6, but does not enter the β-barrel. Also interacts with S30 of the strand 2 and is in line with the chromophore tripeptide.

501	T11C12	5.5	1.1	K107, K126, G127, I128, **K126**	Points toward the space between strands 5 and 6 opposite the N/C termini, but does not enter the β-barrel.
1001	T1C3	7.1	1.5	N159, P187, D190, V193, **G189, G191**	Located in the loop region of strands 9 and 10 and strand 8 (N159), on the same side as the N/C termini.
1001	T2C3	16.7	1.6	K3, A87, P89, E90, P187, G189, D190, G191, V193, **S86, M88, G191**	Some overlap with T1C3, goes in the opposite direction of T1C3 out of the barrel.
1001	T3C4	9.4	1	V22, H25, P54, E132, L137	Interacts with the β-turns of β(1-2), β(6-7), and the loop region of the α-helix (between β3 and helix)
1001	T4C10	5.6	1.5	F99, K101, L178, A179, D180, **F99**	Tunnel points toward space between strands 4 and 9 opposite N/C termini, but does not enter the barrel.
1001	T5C12	4.2	1.9	V11, E34, K41, T43	Hovers over G35. Points right into the space between strands 2 and 3, but does not enter the β-barrel.

Table 16. MOLEonline's predicted channels and their dimensions for the 2AWJ G35A simulation, courtesy of Justin Nwafor**.**

Frame	**Name**	**Length (Å)**	**Bottleneck (Å)**	**Lining Residues**	**Description**
1	T1C1	12.8	1.3	K3, E5, K79, D82, K85, S86, L194, **G4, E5, K79**	Located right next to the N-terminus and the loop between the α-helix and β4. Most of the tunnel is parallel with the bottom of the barrel.
1	T2C1	13.6	1.3	E5, T9, A37, T38, Y74, K79, D82, K85, **F8, A37, Y74**	Some overlap with T1C1, and go out the opposite side of the barrel. Both T2C1 and T1C1 combined to have a horseshoe shape around the

					N-terminus.
1	T3C2	16.5	1.2	P58, Y143, Y145, N146, I167, R168, H169, N170, V176, L207, **N144, R168, N170**	This tunnel enters between β7 and β8, and penetrates the barrel directly to P58, at the top of the helix.
1	T4C3	5.5	3.5	K52, H139, K209, D216	Very large tunnel that interacts with lids regions of the side opposite to the N/C termini.
1	T5C5	22.3	0.7	Y74, F83, F84, I152, M153, A154, I161, L195, P196, D197, N198, H199, **F83, A154, P196, D197, N198**	Long tunnel that enters between β7 and β10 on the side of the N/C termini. Penetrates into the barrel, interacting with the loop after the helix.
1	T6C7	5.5	N/A	T97, F99, Y182	Points towards space between β4 and β9, but does not come close to the barrel at all.
1	T7C8	3.6	2.2	A87, P89, E90, G189, G191, P192, **S86, P192**	Located between the B-termini of β9 and β10 and the loop of the helix on the same side of the N/C termini.
1	T8C9	15.8	0.6	L7, T9, G10, A35, A37, F71, D117, **L7, T9, G10, A35, A37**	This tunnel goes right into the hydrophobic pocket that G35 us typically in.
1	T9C10	16.9	0.9	K101, D102, D103, N135, I136, L141, I171, S175, Q177, **K101, V176**	Interacts with β9, the β4 and β5 turn, and lids opposite to the N/C termini.
1	T10C11	8.4	0.9	K156, N159, V193, L195, **V193**	Located between the loops of β7/β8 and β9/β10 on the same side of the N/C termini. Tunnel points straight up into the barrel, but the tunnel is very much short.
1	T11C12	6	1.2	E111, K113, R122, **V120**	Pointing at the space between β5/β6 (supposed to be strands but it is a loop), on the same

					side as N/C termini.
501	T1C1	8.7	1.9	E5, T38, Y74, K79, K85, **A37, Y74**	Located on the same side as N/C termini, interacts with residues on the β-turn of β2/β3. The loop following the α-helix, and loop of the N-terminus.
501	T2C3	9.6	0.8	V22, H25, K52, P54, V55, L137, **V22, L137**	Located on the turn of β1/β2, interacting with the loop prior to α-helix (opposite to N/C termini) and the loop between β6/β7.
501	T3C4	3.5	1.8	E32, K45, I47, E213	This tunnel is lived by residues that point outward towards bulk solvent, on the 2nd and 3rd β-sheets and loop between β10/β11 (opposite to N/C termini).
501	T4C5	7.4	1.2	Y39, R73, Q204, F223, T225, **Y39, G40, V224**	Funnel shaped tunnel that points into the space between β3/β11. The same side as the N/C termini, but does not go far unto the β-tunnel.
501	T5C7	3.8	2.1	V11, E34, A35, D36, K41, T43	This tunnel points right into the space between β2/β3 where our G35A simulation is, but it does not go into the β-barrel.
1001	T1C1	10.9	1.8	F8, A37, T38, R73, P75, K79, K85, **F8, A37, Y74, D76**	Located near the β-turn of strands 2,3, and the loop region, immediately following the α-helix, and some of the loop following the N-terminus. Runs somewhat parallel with the bottom of the protein.
1001	T2C1	13	1.8	E5, E6, F8, T9, A37, T38, R73, K79, K85, **F8, A37**	Overlaps with T1C1 (fairly perpendicular to each other), points more towards the N-terminus on the way out towards the bulk solvent.

1001	T3C1	19.5	1	E5, T38, R73, Y74, P75, K79, D82, K85, **A37, Y74, D76, K79, H81**	In the same place T1C1 and T2C1, but there's more overlap with this tunnel and T1C1. Narrower size and longer, having more interactions with residues of the post- α-helix loops regions.
1001	T4C2	10.2	1.3	E142, N144, N146, R168, N170, **R168**	Horseshoe shaped with both ends pointing into the space between β7 and β8 (and the loop that follows them), opposite to the N/C termini.
1001	T5C2	13.3	1	P58, V61, Y143, Y145, I167, H169, L207, S208, M218, **Y143, Y145, Y146**	Another horseshoe type tunnel near T4C2, but this tunnel is inside the barrel. One end points out towards the same space that T4C2 points into, and the other points out to the space between β7 and β10 opposite to the N/C termini.
1001	T6C3	9.5	1.8	K3, A87, E90, G191, V193, **G4, S86, P192**	Located right next to the N-terminus and the loop residues between β9 and β10.
1001	T7C3	10.1	1.4	A87, E90, P187, G189, P190, G191, V193, **G189, G191**	Overlap with T6C3, go in the same direction, These tunnels are basically stacked on top of each other.
1001	T8C6	6.8	1.4	K52, L53, W57, H139, Y143, D216	This tunnel interacts with residues near the top of the α-helix (W57). Does not go for enough to interact with chromophore tripeptide.
1001	T9C8	5	1.5	H25, F27, T50, L53, P54, **K26, T50, K52**	Interacts with β2, β3, and the loop prior to α-helix (opposite N/C termini). Points right in the β-barrel through the vertical axis.
1001	T10C12	3.4	2.2	E95, K158, Q184, T186	Tunnel points into space between β4/β9, right under R96, but it does not go into

					the β-barrel.

*Bold lining residues indicate interaction with the backbone of the named residue.

Tight Turn distances:

Chromophore formation occurs through a catalytic autocyclization of Ser65-Tyr66-Gly67. It is necessary for the protein to adopt the "tight turn" conformation for chromophore maturation to occur since this conformation allows the Gly67's amide to attack the Ser65's carbonyl carbon, initiating the maturation process.[29]

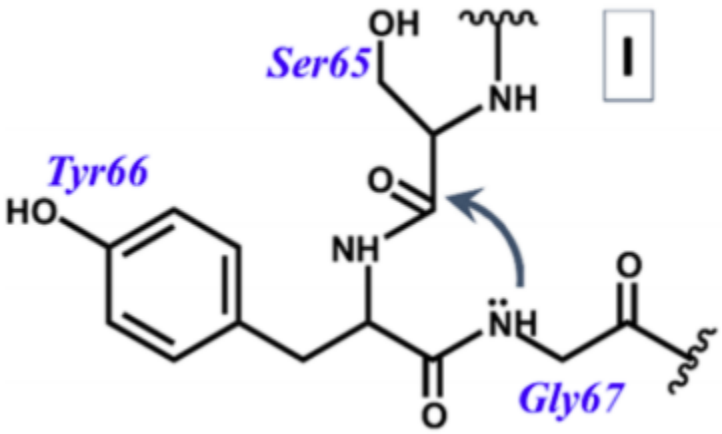


Figure 22. The tight turn conformation results in a short distance between the nitrogen of Gly67 and the carbonyl oxygen that is present in Ser65. For chromophore formation has to be short enough for a nucleophilic attack to proceed.[1]

This distance was tracked through the entirety of the simulations for all structures. It was observed that all the mutants had an increase in the tight turn distance, but it was not enough to make chromophore maturation less likely to happen. Since all the mutations did not have a significant effect, it is unlikely that the glycines have an impact on this distance and chromophore formation.

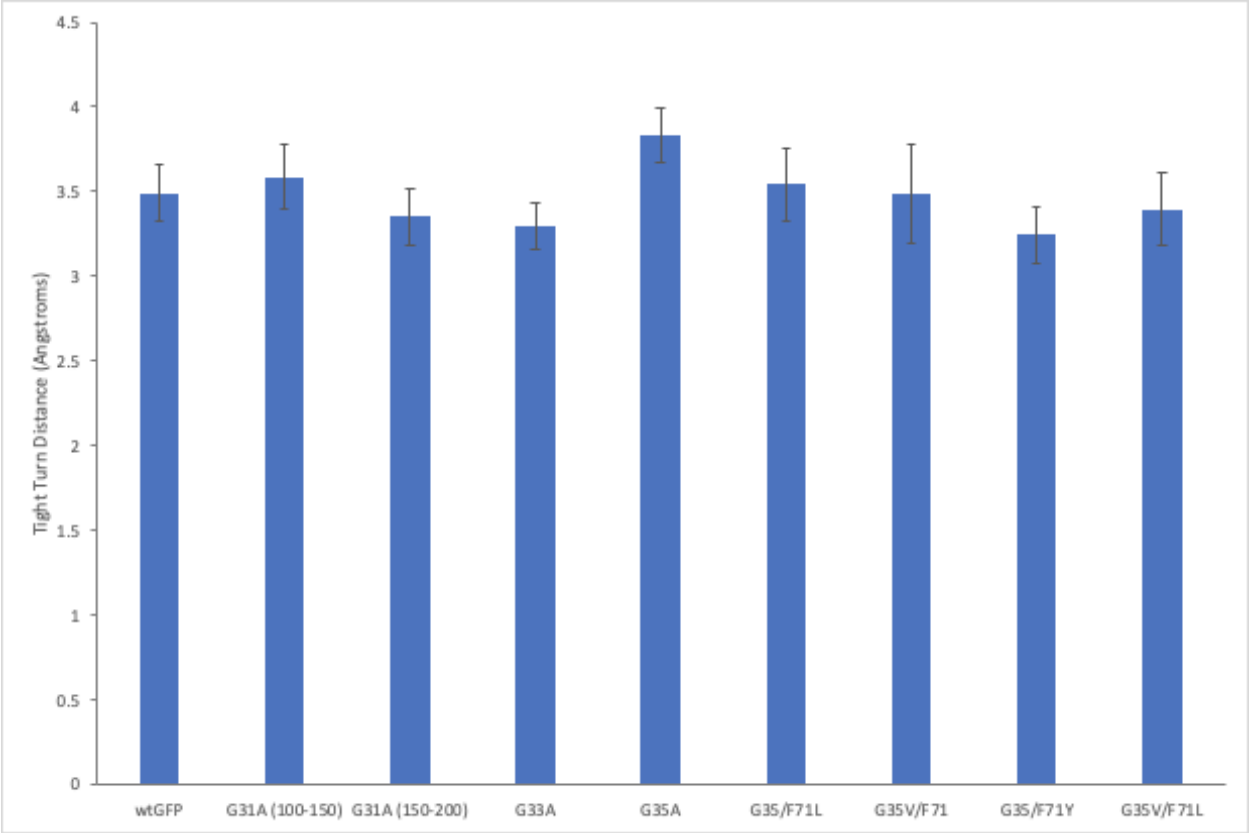


Figure 23. Graph representing the tight-turn distance for all the structures, wtGFP and all the mutants.

Conclusion

Previous analysis of fluorescent proteins (GFP-like) structures revealed that conserved residues fell within 3 categories: chromophore forming, "lid" residues (which serve as hinges) and those located centrally with unknown functions. Within the second β-sheet, there are three highly conserved glycines at positions 31, 33, and 35, with 100%, 87%, and 95% conservation, respectively. These glycines are not observed to be involved in the chromophore formation or are part of the chromophore forming pore. Due to the glycines in the GXGXGX motif being in alternating positions, all three of the glycines are directed towards the protein's core, possibly retarding or hindering the chromophore maturation. To explore this conservation, a number of simulations were run on a modified 2AWJ structure, and its G3XA and hydrophobic pocket mutants (G35/F71L, G35/F71Y, G35V/F71, and G35V/F71L). It was originally hypothesized that the glycines were involved in the chromophore formation considering their proximity to it. The small size and flexible nature of glycines, led us to monitor the distances between the chromophore forming residues, specifically the tight-turn distance. The tight turn conformation increased in all the G3XA mutants, though it was negligible and did not affect the maturation of the chromophore. Thus, it is now suspected that their functionality arises in the protein's folding pathways. A number of key distances were measured throughout the study including H-bonds between strands and aromatic rescue interactions, in order to explore the structural rigidity of the mutants and possible mutational effects. Key α-helical H-bonds were measured for all the G3XA mutants, which were observed to have no significant change, thus no structural change occurred.

An increased distance between strands was observed in all the G3XA mutants, thus an increase of water channels within the β-barrel was expected.

Water migration analyses revealed that there was an increase in water movement around the lids of the residues and a decrease in the α-helical area. This was likely caused by the increased steric effects caused by the methyl groups of the alanines preventing waters from reaching the chromophore in the α-helix area through the strands. The concentrated water movement in the lid area was expected due to the increased bulk-solvent exposure that this area has. In continuation with the observation that G3XA mutants are more likely to misfold and aggregate, it was observed that the β-strands that form the N-terminus were more separated than those found in the wtGFP structure.

One difference that was noted between the mutants and the wtGFP was a larger spacing between G35 and F71, possibly leading to the formation of water channels within the β-barrel. Their interactions were explored through the hydrophobic pocket mutations, which indicated a decrease in the β-barrels rigidity. This was suggested by the observed increase in the H-bond distances between the strands, and expected due to the increased steric effects in the mutated regions. The interaction between the Phe71 and Gly35 was also quantified for all the hydrophobic pocket mutants to determine the effects of the mutations on previously observed aromatic rescue interactions. It was observed that this H-π interaction between these two residues promotes proper packing of the strands.

From these analyses, we were able to determine that the glycines in the positions 31, 33, and 35 do not influence chromophore formation. It is suspected that these three residues and GXGXGX motif are crucial in the proper formation of the β-barrel. Other studies have expanded on the folding pathway intermediates and determined that the β-strands unfold in groups. The

first groups to unfold are the N/C termini strands. The remaining strands move in groups consisting of: β1–3, β4–6 and β7–11.[18]

Our simulations demonstrated diminished H-bonding throughout the β-sheets, thus it is more than likely that the functionality of this triad of glycines is in the folding pathways, as well as to provide proper spacing between the sheets. These findings correlate to those found by Professor Schneider, since the G3XA mutants had an decreased fluorescence and a high propensity to misfolding, thus supporting the importance of the triad within the folding process. Future studies could involve partial structures focusing on β-strands 1-3. The main focus of these would be the folding pathways since these strands remain intact and next to each other for many maturation pathways. It is suspected that these strands could be behaving in a "zipper"-like fashion due to their innate flexibility, which is needed to complete the β-barrel structure.

References

1. Marc Zimmer Green Fluorescent Protein (GFP): Applications, Structure, and Related Photophysical Behavior. *Chemical Reviews.* **2002***, 102*, 759-782.

2. Zimmer, M. GFP: From jellyfish to the Nobel prize and beyond. *Chemical Society reviews* **2009***, 38*, 2823-2832.

3. Matz, M. V.; Markelov, M. L.; Labas, Y. A.; Savitsky, A. P.; Lukyanov, S. A.; Fradkov, A. F.; Zaraisky, A. G. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnology.* **1999***, 17*, 969-973.

4. Zimmer, M. H.; Zimmer, M.; Li, B.; Shahid, R.; Peshkepija, P. Structural consequences of chromophore formation and exploration of conserved lid residues amongst naturally occurring fluorescent proteins. *Chemical Physics.* **2014***, 429*, 5-11.

5. Tsien, R. Y. THE GREEN FLUORESCENT PROTEIN. *Annual review of biochemistry* **1998***, 67*, 509-544.

6. Pédelacq, J.; Tran, T.; Terwilliger, T. C.; Waldo, G. S.; Cabantous, S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology.* **2006***, 24*, 79-88.

7. Remington, S. J. Green fluorescent protein: A perspective. *Protein Science.* **2011***, 20*, 1509-1519.

8. Daria M Shcherbakova; Mikhail Baloban; Alexander V Emelyanov; Michael Brenowitz; Peng Guo; Vladislav V Verkhusha Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nature Communications.* **2016***, 7*, 12405.

9. Daria M Shcherbakova; Vladislav V Verkhusha Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nature Methods.* **2013***, 10*, 751-754.

10. Gerdes, H.; Kaether, C. Green fluorescent protein: applications in cell biology. *FEBS Letters.* **1996***, 389*, 44-47.

11. Chiu, W.; Niwa, Y.; Zeng, W.; Hirano, T.; Kobayashi, H.; Sheen, J. Engineered GFP as a vital reporter in plants. *Current Biology.* **1996***, 6*, 325-330.

12. Stepanenko, O. V.; Stepanenko, O. V.; Kuznetsova, I. M.; Verkhusha, V. V.; Turoverov, K. K. Beta-barrel scaffold of fluorescent proteins: folding, stability and role in chromophore formation. *International review of cell and molecular biology.* **2013***, 302*, 221.

13. Banerjee, S.; Schenkelberg, C. D.; Jordan, T. B.; Reimertz, J. M.; Crone, E. E.; Crone, D. E.; Bystroff, C. Mispacking and the Fitness Landscape of the Green Fluorescent Protein Chromophore Milieu. *Biochemistry.* **2017***, 56*, 736-747.

14. Fu, J. L.; Kanno, T.; Liang, S.; Matzke, A. J. M.; Matzke, M. GFP Loss-of-Function Mutations in Arabidopsis thaliana. *G3 (Bethesda, Md.).* **2015***, 5*, 1849-1855.

15. Ong, W. J.; Alvarez, S.; Leroux, I. E.; Shahid, R. S.; Samma, A. A.; Peshkepija, P.; Morgan, A. L.; Mulcahy, S.; Zimmer, M. Function and structure of GFP-like proteins in the protein data bank. *Molecular bioSystems* **2011***, 7*, 984-992.

16. Dou, J.; Vorobieva, A. A.; Sheffler, W.; Doyle, L. A.; Park, H.; Bick, M. J.; Mao, B.; Foight, G. W.; Lee, M. Y.; Gagnon, L. A.; Carter, L.; Sankaran, B.; Ovchinnikov, S.; Marcos, E.; Huang, P.; Vaughan, J. C.; Stoddard, B. L.; Baker, D. De novo design of a fluorescence-activating β-barrel. *Nature.* **2018***, 561*, 485-491.

17. Sarkisyan, K. S.; Bolotin, D. A.; Meer, M. V.; Usmanova, D. R.; Mishin, A. S.; Sharonov, G. V.; Ivankov, D. N.; Bozhanova, N. G.; Baranov, M. S.; Soylemez, O.; Bogatyreva, N. S.; Vlasov, P. K.; Egorov, E. S.; Logacheva, M. D.; Kondrashov, A. S.; Chudakov, D. M.; Putintseva, E. V.; Mamedov, I. Z.; Tawfik, D. S.; Lukyanov, K. A.; Kondrashov, F. A. Local fitness landscape of the green fluorescent protein. *Nature (London)* **2016***, 533*, 397-401.

18. Govardhan Reddy; Zhenxing Liu; D. Thirumalai Denaturant-dependent folding of GFP. *Proceedings of the National Academy of Sciences - PNAS* **2012***, 109*, 17832-17838.

19. Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW (Nov 2007). "Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system". *Nature.* 450 (7166): 56–62. Bibcode:2007Natur.450...56L. doi:10.1038/nature06293. PMID 17972876. S2CID 4402093.

20. Voss-Andreae J (2005). "Protein Sculptures: Life's Building Blocks Inspire Art". Leonardo. 38: 41–45. doi:10.1162/leon.2005.38.1.41. S2CID 57558522.

21. Wongsrikeao P, Saenz D, Rinkoski T, Otoi T, Poeschla E (2011). "Antiviral restriction factor transgenesis in the domestic cat". Nature Methods. 8 (10): 853–9. doi:10.1038/nmeth.1703. PMC 4006694. PMID 21909101.

22. "Glow-In-The Dark NeonMice". Archived from the original on February 14, 2009. Retrieved August 30, 2016.

23. Katjusa Brejc; Titia K. Sixma; Paul A. Kitts; Steven R. Kain; Roger Y. Tsien; Mats Ormo; S. James Remington Structural Basis for Dual Excitation and Photoisomerization of the Aequorea victoria Green Fluorescent Protein. *Proceedings of the National Academy of Sciences - PNAS* **1997***, 94*, 2306-2311.

24. Wood, T. I.; Barondeau, D. P.; Hitomi, C.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. Defining the Role of Arginine 96 in Green Fluorescent Protein Fluorophore Biosynthesis. *Biochemistry (Easton)* **2005***, 44*, 16211-16220.

25. Schrödinger Release 2021-1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021.

26. Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D.

Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw, "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters," *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida*, **2006**, November 11-17

27. .Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc., **1996**, 118, 11225-11236.*

28. Edward Harder, Wolfgang Damm, Jon Maple, Chuanjie Wu, Mark Reboul, Jin Yu Xiang, Lingle Wang, Dmitry Lupyan, Markus K. Dahlgren, Jennifer L. Knight, Joseph W. Kaus, David S. Cerutti, Goran Krilov, William L. Jorgensen, Robert Abel, and Richard A. Friesner. Journal of Chemical Theory and Computation **2016** 12 (1), 281-296. DOI: 10.1021/acs.jctc.5b00864

29. Grigorenko, B. L.; Krylov, A. I.; Nemukhin, A. V. Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. *Journal of the American Chemical Society* **2017***, 139*, 10239-10249.

30. David P. Barondeau; Christopher D. Putnam; Carey J. Kassmann; John A. Tainer; Elizabeth D. Getzoff Mechanism and Energetics of Green Fluorescent Protein Chromophore Synthesis Revealed by Trapped Intermediate Structures. *Proceedings of the National Academy of Sciences of the United States of America* **2003***, 100*, 12111-12116.

31. Merkel, J. S.; Regan, L. Aromatic rescue of glycine in β sheets. *Folding & design* **1998***, 3*, 449-456.

32. Li, B.; Shahid, R.; Peshkepija, P.; Zimmer, M. Water diffusion in and out of the β-barrel of GFP and the fast maturing fluorescent protein, TurboGFP. *Chemical physics* **2012***, 392*, 143-148.

33. Branchini, B. R.; Nemser, A. R.; Zimmer, M. A Computational Analysis of the Unique Protein-Induced Tight Turn That Results in Posttranslational Chromophore Formation in Green Fluorescent Protein. *Journal of the American Chemical Society* **1998***, 120*, 1-6.