6-1-2008

The Role of the Tight-Turn, Broken Hydrogen Bonding, Glu222 and Arg96 in the Post-translational Green Fluorescent Protein Chromophore Formation

Marc Zimmer
Connecticut College, mzim@conncoll.edu

Nathan P. Lemay
Connecticut College

Alicia L. Morgan
Connecticut College

Elizabeth J. Archer
Connecticut College

Luisa A. Dickson
Connecticut College

See next page for additional authors

Recommended Citation

Lemay, N.P.; Morgan, A.L.; Archer, E.J.; Dickson, L.A.; Megley, C.M.; Zimmer, M. The role of the tight-turn, broken hydrogen bonding, Glu222 and Arg96 in the post-translational green fluorescent protein chromophore formation. *Chem. Phys.* **2008**, *348*, 152-160. http://dx.doi.org/10.1016/j.chemphys.2008.02.055.

Published in final edited form as:
Chem Phys. 2008 June 2; 348(1-3): 152–160. doi:10.1016/j.chemphys.2008.02.055.

The Role of the Tight-Turn, Broken Hydrogen Bonding, Glu222 and Arg96 in the Post-translational Green Fluorescent Protein Chromophore Formation

Nathan P. Lemay, **Alicia L. Morgan**, **Elizabeth J. Archer**, **Luisa A. Dickson**, **Colleen M. Megley**, and **Marc Zimmer***
Dept. of Chemistry, Connecticut College, New London, CT 06320

Abstract

Green Fluorescent Proteins (GFP) and GFP-like proteins all undergo an autocatalytic post-translational modification to form a centrally located chromophore. Structural analyses of all the GFP and GFP-like proteins in the protein databank were undertaken to determine the role of the tight-turn, broken hydrogen bonding, Gly67, Glu222 and Arg96 in the biosynthesis of the imidazolone group from 65SYG67. The analysis was supplemented by computational generation of the conformation adopted by uncyclized wild-type GFP. The data analysis suggests that Arg96 interacts with the Tyr66 carbonyl, stabilizing the reduced enolate intermediate that is required for cyclization; the carboxylate of Glu 222 acts as a base facilitating, through a network of two waters, the abstraction of a hydrogen from the α-carbon of Tyr66; a tight-turn conformation is required for autocatalytic cyclization. This conformation is responsible for a partial reduction in the hydrogen bonding network around the chromophore-forming region of the immature protein.

Keywords

GFP; Post-translational modification; Fluorescent proteins; Conquest; Database analysis

INTRODUCTION

In the last ten years green fluorescent protein (GFP) has changed from a nearly unknown protein to a commonly used molecular imaging tool in biology, chemistry, genetics and medicine.(1–8) Probably the best indicator of the utility of GFP and GFP-like proteins is the fact that in 2004 about 50%, 35%, 60% and 20% of the articles in *Cell*, *Development*, *Journal of Cell Biology* and *Neuron* mentioned or used GFP-like proteins(4) and that in 2006 more than 10,000 papers were published that used GFP. GFPs and GFP-like proteins are particularly useful due to their stability and the fact that the chromophore is formed in an autocatalytic cyclization that does not require a cofactor. The GFP chromophore is formed autocatalytically by cyclization of the protein backbone.(9–15) This means that unlike most other bioluminescent reporters, GFP fluoresces in the absence of any other proteins, substrates, or cofactors. Furthermore, it appears that fusion of GFP to a protein does not alter the function or location of the protein.

*mzim@conncoll.edu; fax 860-439-2477; http://www.conncoll.edu/ccacad/zimmer/GFP-ww/GFP-1.htm.

In the intact GFP, the chromophore is formed by an autocatalytic internal cyclization of the tripeptide 65SYG67 and subsequent oxidation of the intrinsically formed structure. GFP fluorescence is not observed until 90 minutes to 4 hours after protein synthesis.(10,16) Apparently, the protein folds quickly but the subsequent fluorophore formation and oxidation is slow.(17) GFP refolding from an acid-, base-, or guanidine HCl-denatured state (chromophore containing but non-fluorescent) occurs with a half-life of between 24 seconds (18) and 5 minutes(19) and the recovered fluorescence is indistinguishable from that of native GFP.(20) In 1994 Tsien(10,21) proposed that the chromophore is formed by nucleophilic attack of the nitrogen of Gly67 on the carbonyl carbon of Ser65, see Scheme 1.

The most obvious conformational prerequisite for the autocatalytic cyclization step, shown in Scheme 1, is that the nitrogen of Gly67 and the carbonyl carbon of Ser65 have to be in close proximity to each other. This conformation which preorganizes Ser65-Gly67 for cyclization is termed the tight-turn conformation.(11) Early calculations have shown that in immature precyclized wild-type GFP the distance between the amide nitrogen of Gly67 and the carbonyl carbon of Ser65 is shorter than the sum of their covalent radii (3.25 Å).(22)

Before the crystal structure of GFP or any of its mutants was solved it was also suggested that a tight-turn conformation in immature GFP was not the only requirement for an autocatalytic cyclization. It was proposed that an arginine residue would be found in close proximity to the chromophore-forming region and that it had a catalytic function in the chromophore formation – namely to activate the carbonyl group of Ser65.(22) The subsequently released crystal structures of GFP(23,24) revealed that an arginine, Arg96, was indeed in the position predicted.

In the last decade three different mechanisms for chromophore formation have been proposed.

1. Getzoff et al.(13) solved the crystal structures of both the mature and immature (precyclized) forms of the R96A GFP mutant, and the mature (aerobic) and immature (anaerobic) structures of the S65G, Y66G GFP mutant.(13) Based on these and other crystal structures they proposed a cyclization-dehydration-oxidation mechanism in which

- Arg96 has both a steric(14) and electronic function in chromophore formation.
- The GFP protein matrix creates a dramatic bend of the central helix at the chromophore that removes specific main-chain hydrogen bonds, which must otherwise be broken during maturation.
- The unfavorable cyclization reaction is driven by a subsequent trapping reaction, most likely the dehydration of the carbonyl oxygen of residue 65, see Figure 1 top(13) and Figure 10.

2. Wachter et al. have proposed the cyclization-oxidation-dehydration mechanism(12,25) see Figure 1 middle. It is based on the crystal structure of a Y66L colorless GFP intermediate, substitution studies of Gly 67, Arg96 and Glu222,(25,26) and a kinetic trace for hydrogen peroxide evolution in chromophore formation.(15,27)

3. Density functional calculations led to the proposal that the cyclization is preceded by the dehydration of Tyr66 to dehydrotyrosine, see Figure 1 bottom.(28) In this study two mechanisms were analyzed, the commonly accepted mechanism in which cyclization precedes dehydrogenation (reduced mechanism) and a mechanism in which dehydrogenation precedes cyclization (oxidized mechanism). The reduced pathway was not supported by the calculations, while the alternative oxidized mechanism where a dehydration occurs prior to the formation of the ring yields reasonable energetics for the system. These calculations were done before the Wachter and Getzoff mechanisms were proposed and the energetics of conjugate trapping were not considered. In these mechanisms the initial cyclization reaction is thermodynamically

unfavorable, as shown by our DFT calculations, but the reaction is driven by the chromophore maturation, which we had not considered.

We have used a database analysis of all the crystal structures of GFP and GFP-like proteins in the protein databank (pdb)(29) and a computational analysis of immature wild-type GFP to examine the role of the tight-turn, broken hydrogen bonding, Glu222 and Arg96 in GFP chromophore formation.

MATERIALS AND METHODS

The coordinates of the R96M GFP mutant(30) (2AWJ) were obtained from the Protein Data Bank.(29) Maestro v7.0113 was used to add hydrogen atoms where needed, and the methionine residue in position 96 was graphically converted back to an arginine residue.

The AMBER* force field of MacroModel v9.0016(31) was used in all calculations. A “hot” area with a radius of 8Å (complete residues) from the chromophore (residues 65–67) and Arg96 was used. It was held in place with two subsequent sub-shells each extending an additional 2.00Å with increasing atomic restraints of 100kJ/Å and 200kJ/Å. The Polak-Ribiere conjugate gradient minimization mode was used with a derivative convergence criterion of 0.05kJ/mol. Conformational searches were conducted using the mixed Monte Carlo torsional and molecular position variation method/large scale low mode searching.(32–35) The flexible dihedral angles of all the side-chains of residues that are within 6.00 Å of the chromophore and Arg96 were randomly rotated by between 0 and 180°, and all solvent molecules in that sphere were randomly rotated and translated by between 0 and 1.00Å in each Monte Carlo (MC) step. (36) The main chain dihedrals of residues 62–69 were varied and the amide bonds between residues 60 and 61, and 70 and 71 were designated as closure bonds with minimum and maximum closure distances of 1 Å and 5 Å respectively. 65,000 MC steps were undertaken during the search. Structures within 50kJ/mol of the lowest energy minimum were kept and a usage directed method(33) was used to select structures for subsequent MC steps. Structures found in the conformational search were considered unique if the least squared superimposition of equivalent non-hydrogen atoms found one or more pairs separated by 0.25Å or more. The lowest energy structure obtained in the Monte Carlo torsional and molecular position variation search was further subjected to a 35,000 step large scale low mode conformational search. (34,35)

An in-house database of all the GFP and GFP-like proteins in the Protein Data Bank(29) as of June 1, 2007 was created for use with Conquest.(37) The database contains 70 cyclized GFPs, 8 immature GFPs and 34 GFP-like structures, see supplementary material. The geometrical properties of these structures were analyzed with Conquest,(37) Vista and Isostar.(38)

RESULTS AND DISCUSSION

Arg96 catalyzes chromophore formation, but it is not essential as seen in the R96C mutant which is fluorescent.(39) This is further supported by Getzoff et al.(13) who have shown that the R96A mutation slows the cyclization reaction from minutes to months. In 2005 they also published the crystal structure of the R96M mutant.(30) This mutant is very interesting to us as the methionine sidechain is similar in volume to that of the catalytically important arginine sidechain, but lacks the positive charge of arginine’s guanidinium group. The crystal structure of the mutant has the chromophore in the immature form as it does not catalyze chromophore cyclization. We have used this structure to computationally generate the structure of the immature wild-type GFP because it is the R96M mutant that is structurally closest to the precyclized form of wild-type GFP. A thorough conformational search was performed using

the graphically generated immature wild-type GFP and the resultant low energy conformations were used to examine chromophore formation in GFP.

The conformational search of the immature form of wild-type GFP found 10,606 structures within 50kJ/mol of the lowest energy structure. These structures were analyzed to establish if the immature form of wild-type GFP has a tight turn conformation, whether the GFP protein matrix creates a dramatic bend at the chromophore of the central helix that removes the main-chain hydrogen bonds in the chromophore forming region, and what the role of the Arg96 and Glu222 are in the chromophore formation. They were also compared to all the GFP and GFP-like proteins in the Protein Data Bank.(29) At the time this research was done the pdb contained 70 cyclized GFPs, 8 immature GFPs and 34 GFP-like structures.

Tight-Turn

In order for the nitrogen of Gly67 to attack the carbonyl carbon of Ser65 they have to be in close proximity to each other. It has been suggested that GFP in its immature form accomplishes this by adopting the tight-turn conformation.(40) In the tight-turn conformation the distance between the two reacting atoms is shorter than the sum of their covalent radii (3.25 Å). The distance between the amide nitrogen of Gly67 and the carbonyl carbon of Ser65 in the lowest energy conformation of the calculated immature wild-type GFP structures is 3.13 Å and the average of all the lowest energy structures obtained is 3.12 Å (min 3.08 Å, max. 3.15 Å, SD 0.006 Å). Table I and Figure 2 show that the calculated immature conformations of wild-type GFP, as well as all the GFP structures with an immature chromophore in the pdb have a tight-turn conformation. The calculated immature form differs from the solid state immature forms, this is most likely due to the hydrogen bond formed between the carbonyl oxygen of Tyr66 and Arg96 (which is not present in most immature GFPs, see below).

Glycine67

All reported mutations of Gly67 in GFP and GFP-like proteins have produced non-fluorescent proteins.(39,41) Molecular modeling studies have shown that Gly67 is crucial to adoption of the tight-turn conformation.(11) Therefore it is not surprising that all GFP and GFP-like proteins with mature chromophores in the pdb have a Gly67 (GFP numbering). The ψ and φ dihedrals of Gly67 in all the immature crystal structures are in the favored regions of a glycine Ramachandran plot. The calculated structures are in an allowed region. While the ψ and φ dihedrals of Gly67 in all the immature crystal structures are in the dihedral space that favors α-helical formation the lowest energy immature GFP conformations calculated as described in this paper are outside the α-helical region.(42)

Hydrogen Bonding

The crystal structures of the immature GFP mutants have shown that the GFP protein matrix creates a dramatic bend at the chromophore of the central helix.(13) This kink removes main-chain hydrogen bonds that are commonly associated with an alpha helix. It is presumed that this aids in chromophore formation because the hydrogen bonds would otherwise be broken during maturation. An examination of all the low energy conformational families showed that they all have a kink in their central helix and a disrupted hydrogen bonding pattern around the chromophore forming region, see Figure 3 and Table II.

The calculated immature wild-type structure differs from the crystal structures of the GFP mutants that have been arrested in the immature form, see Figure 2 and Table III. The main reason for this is that the calculated immature structure forms a hydrogen bond between the carbonyl of Ser65 and the amide hydrogen of Gly67, which is not present in the immature GFP structures found in the pdb.

In order to establish the role of the tight turn in disrupting the hydrogen bonding pattern normally found in an ideal alpha helix, we constructed a decapeptide VTTFSYGVQC (corresponding to GFP residues 61 to 70) in an ideal alpha helical conformation and minimized it with a 3.00 Å constraint (k = 100kJ/Å) between the nitrogen of Gly67 and the carbonyl carbon of Ser65. The constraint, which enforced a tight-turn conformation, resulted in the disruption of the Ser65 (C=O) to Val68 (N-H) hydrogen bonding interaction. Thus, the tight-turn is responsible for breaking up some of the canonical hydrogen bonds in an ideal α-helix.

The Role of Arg96 in Chromophore Formation

Although Arg96 is not absolutely necessary for chromophore formation, it is highly conserved in GFPs.(43) In the pdb there are 103 structures of GFP or GFP-like proteins in which cyclization has occurred. All these structures, except 2AWK(30) (R96M), 2AWL(30) (R96K) , 2AWM(30) (R96A) and 1QYF(44) (R96A) have an arginine in a position equivalent to that of Arg96 in wild-type GFP. An Isostar plot, see Figure 4, of the 99 structures containing Arg96 or an Arg in an equivalent position shows that all the arginines are in the same area relative to the cyclized five-membered ring. Since all the mature GFPs, except the 4 slow-maturing GFPs, have an arginine in the same position, it is safe to assume that Arg96 plays an important role in catalyzing the cyclization of the GFP chromophore.

Figure 5A shows the distribution of the angles between the best plane of the imidazolone ring and the guanidinium group of all the arginines shown in Figure 5. Figure 5B shows the distribution of the distances between the carbonyl carbon of the imidazolone and the closest hydrogen of the arginines. The histograms confirm the fact that Arg96 or its equivalents are all located in the same space relative to the chromophore.

Three different roles for Arg96 in the autocatalytic cyclization have been proposed, see Figure 6. The orientation of Arg96 relative to the imidazolone ring should help in ruling out some of these proposals. Analysis of the 10,606 GFP structures with the immature form of the chromophore reveals that none of the structures have a distance of less than 5.00 Å between the carbonyl oxygen of Ser65 and any of the hydrogens on Arg96 (Max. = 5.64 Å; Min. = 5.24 Å; Mean = 5.39 Å). This seems to eliminate the Zimmer ’98 proposal (see Figure 6).(22) On the other hand the distance from Arg96 to the carbonyl oxygen of residues 66 is much shorter (Max. = 2.02 Å; Min. = 1.75 Å; Mean = 1.83 Å) leading us to conclude that our initial proposed role for Arg96 is incorrect and that chromophore cyclization is probably catalyzed in the manner proposed by Getzoff et al.(13) and modified by Wachter,(15) see Figure 7 left.

The Role of Glu222 in Chromophore Formation

Like Arg96, Glu222 is not absolutely necessary for chromophore formation and variants such as the E222G mutant fluoresce.(45) However, since Wachter et al. have proposed a role for Glu222 in chromophore formation(25) (see Figure 7) and Glu222 is highly conserved in GFPs, (43) we have examined the geometry and structural environment of Glu222.

Based on kinetic and mass spectroscopy experiments conducted at different pHs it has been suggested that Glu222 can act as a base facilitating proton abstraction from the Gly67 nitrogen or the Tyr66 α-carbon, see Figure 7. In the pdb there are 103 structures of GFP or GFP-like proteins in which cyclization has occurred. All these structures, except 1W7T,(46) 1W7U, (46) 2HGY,(47) and 1HCJ,(48) have a glutamic acid in a position equivalent to that of Glu222 in wild-type GFP. 1W7T,(46) 1W7U(46)and 1HCJ(48) have incomplete Glu222 side-chains due to photochemical decarboxylation, and 2HGY is a E222A mutant. An isostar plot, see Figure 5, of the 99 structures containing Glu222 or a Glu in an equivalent position shows that all the Glu sidechains are in the same area relative to the cyclized five-membered ring. However, they are not as concentrated in space as the Arg96 residues. Based on the fact that

Glu222 abstracts a hydrogen from either the Gly67 nitrogen or the Tyr66 α-carbon, we examined the possible interactions between Glu222 and Gly67 or Tyr66. The distances between the nitrogen of Gly67 and the closest oxygen of Glu222 vary between 3.949Å and 7.197 Å (Mean 5.806 Å, SD 0.72 Å), the distances between the α-carbon and the closest oxygen of Glu222 vary between 3.537 Å and 6.069 Å (Mean 4.854 Å, SD 0.63 Å), and the angles between the plane of the imidazolone rings and the Glu222 carboxyl group vary from 2.73° to 12.93° (Mean 7.66°, SD 2.783°). Glu222 was found in a similar position in the calculated immature wild-type structures, but the Glu222 was a little further from the 65SYG67 tripeptide than observed in the cyclized GFP crystal structures. The distances between the nitrogen of Gly67 and the closest oxygen of Glu222 vary between 7.60Å and 9.21 Å (Mean 7.60 Å), the distances between the α-carbon and the closest oxygen of Glu222 vary between 5.59 Å and 7.17 Å (Mean 5.88 Å),

The distances listed above are too long for Glu222 to act directly as a base, therefore if it is to act as a base, then hydrogen transfer involving water molecules will be required. We examined the lowest energy immature wild-type GFP structure and found a hydrogen bonding network such as that required for hydrogen abstraction from the α- carbon of Tyr66, Figure 7 right. The hydrogen transfer pathway, Figure 8, is robust and is found in most of the 10,606 structures.

For the 10,606 low energy conformations the average distance between carbonyl oxygen of Glu222 and the hydrogen of water1042 (waters numbered according to 2AWJ) is 1.75 Å, the average distance between the waters is 1.83 Å, and the average distance between the oxygen of water1158 and the hydrogen of the α-carbon of Tyr66 is 2.85 Å. Therefore Glu222 can act as a base activating the nitrogen of Gly67 for nucleophilic addition to the carbonyl carbon of Ser65, see Figure 7 right, or it can be responsible for the abstraction of the hydrogen on the α-carbon of Tyr66, facilitating the enolate formation shown in Figure 9.(14) No similar hydrogen bonding network was found between Glu222 and the amide nitrogen of Gly67.

Although histidine ammonia lyase (HAL),(49,50) phenylalanine ammonia lyase (PAL),(51) and tyrosine aminomutase (TAM)(52) undergo a similar autocatalytic post-translational cyclization to that occurring in GFP, Table III shows that there is no arginine in a position equivalent to that of Arg96 in GFP and that if it has any role, arginine clearly does not have the same role in the autocatalytic cyclization that it does in GFP. The distances between the closest glutamic acid residue and the central α-carbon of the cyclized tripeptide are similar to those found in GFP. However, the glutamic acid residues are located in different areas relative to the imidazolone ring to the Glu222s in GFP, see Isostar plot in Figure 10, and there is no hydrogen transfer chain such as that shown in Figure 8. We therefore suggest that the post-translational modification in HAL, PAL and TAM does not involve a glutamic acid acting as a general base.

CONCLUSIONS

All the immature structures examined, that is both the crystal and calculated structures, have a tight-turn conformation in which the nitrogen of Gly67 is closer to the carbonyl carbon of Ser65 than the sum of their covalent radii. This tight turn conformation is responsible for a kink in the chromophore-forming region of the α-helix. The kink results in a reduction in the number of hydrogen bond interactions in the 65SYG67 tripeptide fragment, which might aid in the formation of the chromophore as chromophore biosynthesis will not involve energetically unfavorable hydrogen bond breakage.(13) The calculated structures of immature wild-type GFP have a hydrogen bond between Ser65(**C**=O) and Gly67(**N**) that is not found in the solid state structures. Gly67 is found in all GFP and GFP-like structures. Our earlier calculations(11) show that it is required for tight-turn formation.

Figure 5 shows that the guanidinium group of Arg96 (GFP numbering) is located in the same region of space relative to the imidazolone ring in all GFP and GFP-like structures. The distance between the closest Arg96 hydrogen and the carbonyl carbon of Gly is shorter than 2.25 Å in all structures examined. We conclude that in GFPs Arg96 catalyses the post-translational ring formation by favoring Gly67 deprotonation through electrostatic interactions with the Tyr66 carbonyl oxygen and by stabilizing the enolate form. This role for Arg96 in chromophore formation is a feature of the both the Getzoff and Wachter mechanisms.(14,25) Table III shows that the post-translational cyclization in HAL, PAL and TAM cannot involve an arginine stabilized enolate form.

In all Glu222s GFP and GFP-like proteins are located in the same area, but are too far from the 65SYG67 tripeptide to directly abstract a proton. If Glu222 acts as a base it has to do this through a network of waters. The calculated structures have two waters in the correct position for abstraction of a hydrogen from the α-carbon of Tyr66, see Figure 9. Glu222 most likely acts as a base removing the proton from the α-carbon of Tyr66, which stabilizes the enolate form of Gly67 as shown in Figure 7 right or Figure 9. However, post-translational cyclization in HAL, PAL and TAM does not involve a glutamic acid having the same role as Glu222 in GFP.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by NIH grant R15GM059108-03 and the Barbara Zaccheo Kohn ’72 endowment.

REFERENCES

1. Zhang J, Campbell RE, Ting AY, Tsien RY. Nature Reviews Molecular Cell Biology 2002;3:906–918.
2. Zimmer, M. Glowing Genes: A Revolution in Biotechnology. Amherst, N.Y.: Prometheus Books; 2005.
3. Tsien RY. Febs Letters 2005;579:927–932. [PubMed: 15680976]
4. Chalfie, M.; Kain, SR., editors. Green Fluorescent Protein: Properties, Applications and Protocols. Second Ed.. Hoboken, New Jersey: John Wiley & Sons; 2006.
5. Tozzini, V.; Pellegrini, V.; Beltram, F. CRC Handbook of Organic Photochemistry and Photobiology. 2nd Edition. 2004. Green Fluorescent Proteins and Their Applications to Cell Biology and Bioelectronics.
6. Chudakov DM, Lukyanov S, Lukyanov KA. Trends in Biotechnology 2005;23:605–613. [PubMed: 16269193]
7. Shaner NC, Steinbach PA, Tsien RY. Nature Methods 2005;2:905–909. [PubMed: 16299475]
8. Leutenegger A, D'Angelo C, Matz MV, Denzel A, Oswald F, Salih A, Nienhaus GU, Wiedenmann J. Febs Journal 2007;274:2496–2505. [PubMed: 17419724]
9. Cody CW, Prasher DC, Westler WM, Pendergast FG, Ward WW. Biochemistry 1993;32:1212–1218. [PubMed: 8448132]
10. Heim R, Prasher DC, Tsien RY. Proc. Natl. Acad. Sci. USA 1994;91:12501–12504. [PubMed: 7809066]
11. Branchini BR, Nemser AR, Zimmer M. J. Am. Chem. Soc 1998;120:1–6.
12. Rosenow MA, Huffman HA, Phail ME, Wachter RM. Biochemistry 2004;43:4464–4472. [PubMed: 15078092]
13. Barondeau DP, Putnam CD, Kassmann CJ, Tainer JA, Getzoff ED. Proceedings of the National Academy of Sciences of the United States of America 2003;100:12111–12116. [PubMed: 14523232]

14. Barondeau DP, Kassmann CJ, Tainer JA, Getzoff ED. Biochemistry 2005;44:1960–1970. [PubMed: 15697221]
15. Wachter RM. Accounts of Chemical Research 2007;40:120–127. [PubMed: 17309193]
16. Crameri A, Whitehorn EA, Tate E, Stemmer WPC. Nat. Biotechnol 1996;14:315–319. [PubMed: 9630892]
17. Reid BG, Flynn GC. Biochemistry 1997;36:6786–6791. [PubMed: 9184161]
18. Makino Y, Amada K, Taguchi H, Yoshida M. J. Biol. Chem 1997;272:12468–12474. [PubMed: 9139695]
19. Ward WWB, Bokman SH. Biochemistry 1982;21:4535–4540. [PubMed: 6128025]
20. Bokman SH, Ward WW. Biochem. Biophys. Res. Commun 1981;101:1372–1380. [PubMed: 7306136]
21. Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA. RY Tsien TIBS 1995;20:448–455.
22. Zimmer, M.; Branchini, BR.; Lusins, JO. Bioluminescence and chemiluminescence: Proceedings of the 9th International Symposium 1996. Hastings, JW.; Kricka, LJ.; Stanley, PE., editors. Chichester, England: John Wiley; 1996. p. 407-410.
23. Ormoe M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Science 1996;273:1392–1395. [PubMed: 8703075]
24. Yang F, Moss L, Phillips G. Nature Biotech 1996;14:1246–1251.
25. Sniegowski JA, Lappe JW, Patel HN, Huffman HA, Wachter RM. Journal of Biological Chemistry 2005;280:26248–26255. [PubMed: 15888441]
26. Sniegowski JA, Phail ME, Wachter RM. Biochemical and Biophysical Research Communications 2005;332:657–663. [PubMed: 15894286]
27. Zhang LP, Patel HN, Lappe JW, Wachter RM. Journal of the American Chemical Society 2006;128:4766–4772. [PubMed: 16594713]
28. Siegbahn PEM, Wirstam M, Zimmer M. Internat. J. Quantum Chem 2001;81:169–186.
29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. Nucleic Acids Research 2000;28:235–242. [PubMed: 10592235]
30. Wood TI, Barondeau DP, Hitomi C, Kassmann CJ, Tainer JA, Getzoff ED. Biochemistry 2005;44:16211–16220. [PubMed: 16331981]
31. MacroModel, version 9.0. New York, NY: Schroedinger, LLC; 2005.
32. Chang G, Guida WC, Still WC. J. Am. Chem. Soc 1989;111:4379–4386.
33. Saunders M, Houk K, Wu Y-D, Still W, Lipton M, Chang G, Guida W. J. Am. Chem. Soc 1990;112:1419.
34. Kolossvary I, Keseru GM. Journal of Computational Chemistry 2001;22:21–30.
35. Keseru GM, Kolossvary I. Journal of the American Chemical Society 2001;123:12708–12709. [PubMed: 11741448]
36. Bartol J, Comba P, Melter M, Zimmer M. J Comput Chem 1999;20:1549–1558.
37. Bruno IJ, Cole JC, Edgington PR, Kessler M, Macrae CF, McCabe P, Pearson J, Taylor R. Acta Crystallographica Section B-Structural Science 2002;58:389–397.
38. Bruno IJ, Cole JC, Lommerse JPM, Rowland RS, Taylor R, Verdonk ML. Journal of Computer-Aided Molecular Design 1997;11:525–537. [PubMed: 9491345]
39. Tsien RY. Annu. Rev. Biochem 1998;67:509–544. [PubMed: 9759496]
40. Branchini BR, Lusins JO, Zimmer M. J. Biomol. Struct, & Dyn 1997;14:441–448. [PubMed: 9172644]
41. Cormack BP, Valdiva RH, Falkow S. Gene 1996;173:33–38. [PubMed: 8707053]
42. Branden, C.; Tooze, J. Introduction to Protein Structure. New York: Garland Publishing; 1998.
43. Matz MV, Fradkov AF, Labas YA, Savitisky AP, Zaraisky AG, Markelov ML, Lukyanov SA. Nature Biotech 1999;17:969–973.
44. Barondeau DP, Putnam CD, Wood TI, Kassmann CJ, Tainer JA, Getzoff ED. Protein Science 2004;13:153–153.
45. Ehrig T, O'Kane DJ, Pendergast FG. FEBS Letters 1995;367:163–166. [PubMed: 7796912]

46. van Thor JJ, Georgiev GY, Towrie M, Sage JT. Journal of Biological Chemistry 2005;280:33652–33659. [PubMed: 16033764]

47. Barondeau DP, Kassmann CJ, Tainer JA, Getzoff ED. Journal of the American Chemical Society 2007;129:3118–3126. [PubMed: 17326633]

48. van Thor JJ, Gensch T, Hellingwerf KJ, Johnson LN. Nature Structural Biology 2002;9:37–41.

49. Schwede TF, Retey J, Schulz GE. Biochemistry 1999;38:5355–5361. [PubMed: 10220322]

50. Rother R, Poppe L, Viergutz S, Langer B, Retey J. European Journal of Biochemistry 2001;268:6011–6019. [PubMed: 11732994]

51. Calabrese JC, Jordan DB, Boodhoo A, Sariaslani S, Vannelli T. Biochemistry 2004;43:11403–11416. [PubMed: 15350127]

52. Christenson SD, Liu W, Toney MD, Shen B. Journal of the American Chemical Society 2003;125:6062–6063. [PubMed: 12785829]

53. Kabsch W. Acta Crystallographica Section A 1976;32:922–923.

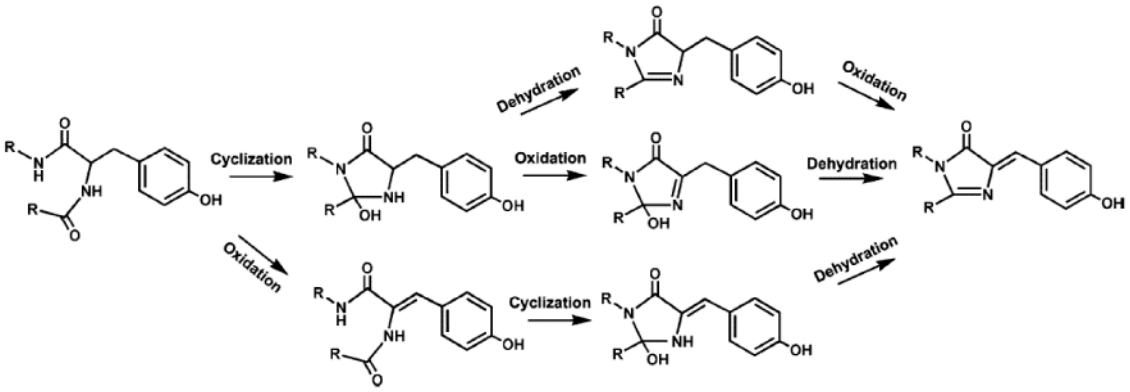


Figure 1.

Three different proposed reaction schemes for chromophore formation. The cyclization-dehydration-oxidation Getzoff mechanism,(13) the cyclization-oxidation-dehydration Wachter mechanism,(12) and the dehydration-cyclization-oxidation Siegbahn mechanism.(28) (courtesy *Biochemistry* 2005, **44**, 1962.)

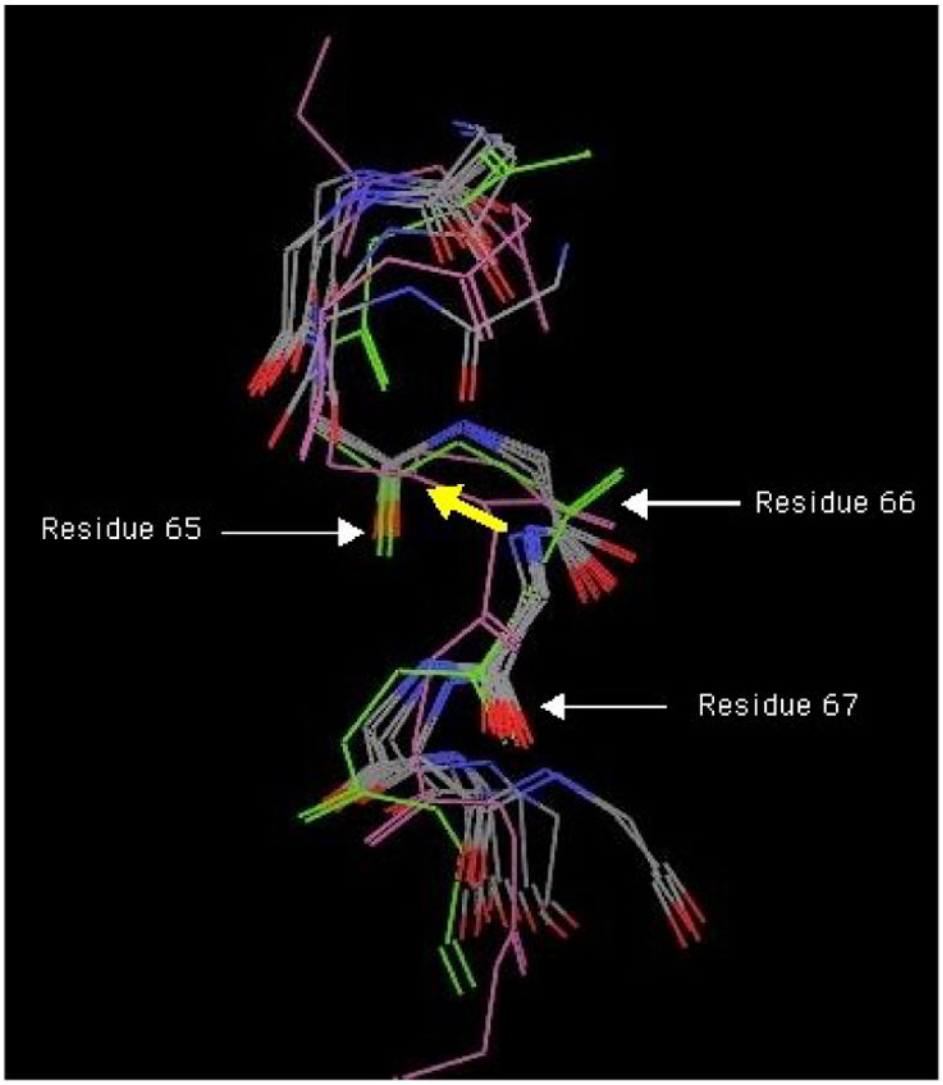


Figure 2.

Least squared superimposition(53) of equivalent non-hydrogen main chain atoms in residues 65–67. Residues 63–69 of all immature GFP mutants in the pdb (1QXT,(44) 1QY3,(44) 1QYO, (13) 1YHG,(14) 1YHH,(14) 1YHI,(14) 2AWJ,(30) 2HFC(47)) are shown colored according to their atom identities, while the lowest energy conformation of immature wild-type GFP is colored green. The yellow arrow shows the short (less than the sum of the van der Waals radii) distance between the nitrogen of residue 67 and the carbonyl carbon of residue 65. The structure of mature wild-type GFP (1GFL(24)) was added in pink for comparison purposes.

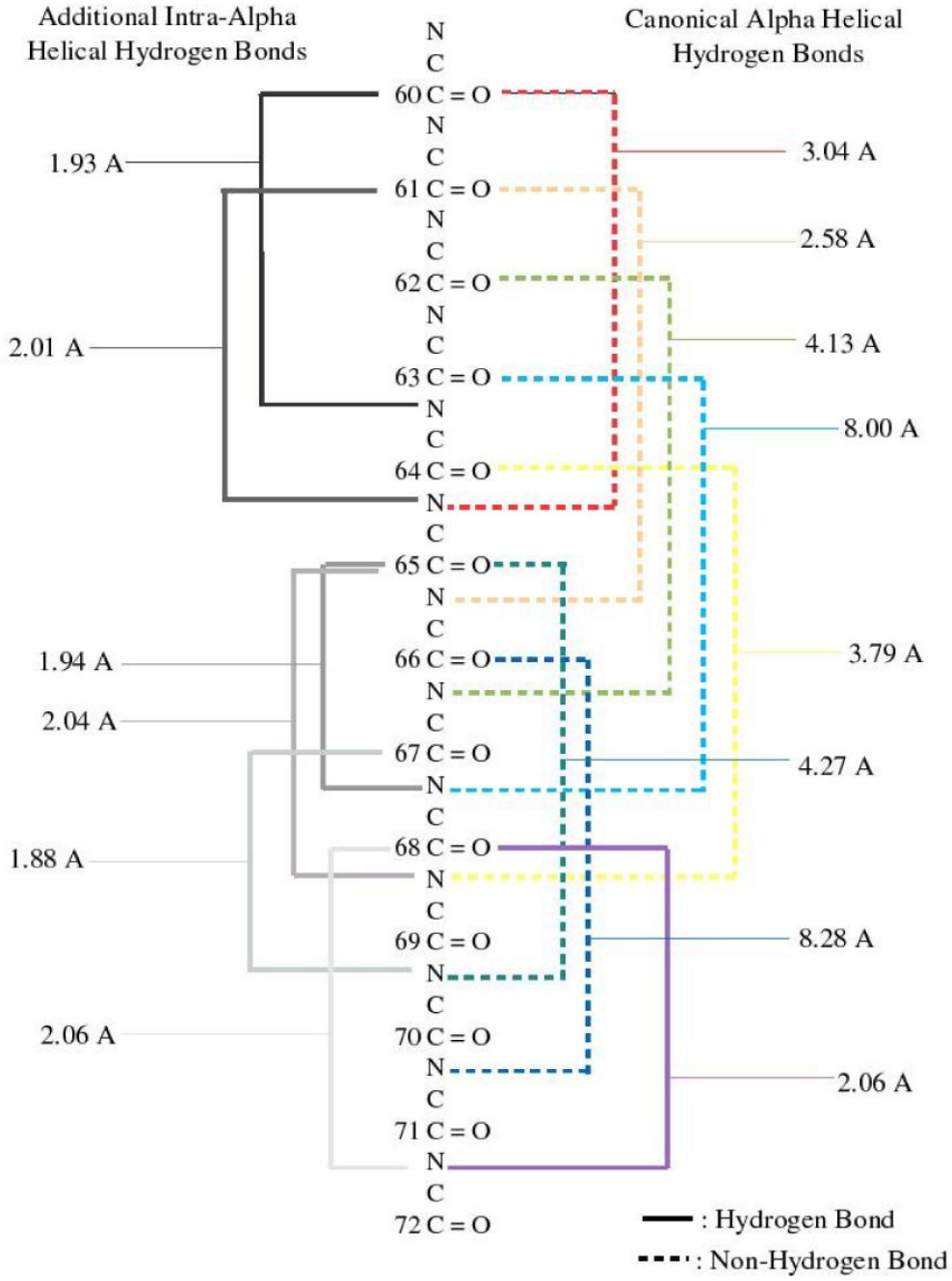


Figure 3.
Main-chain hydrogen bonding pattern found in the lowest energy conformation of the computationally generated immature wild-type GFP. Solid lines indicate non-bonded interactions that meet the Maestro requirements set for hydrogen bonds.

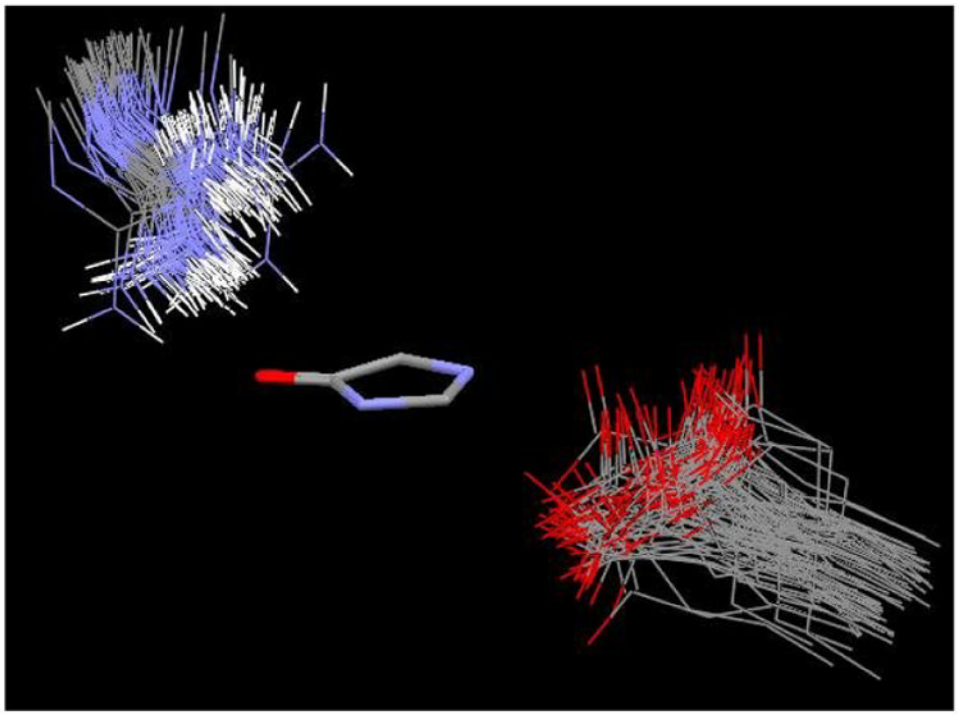
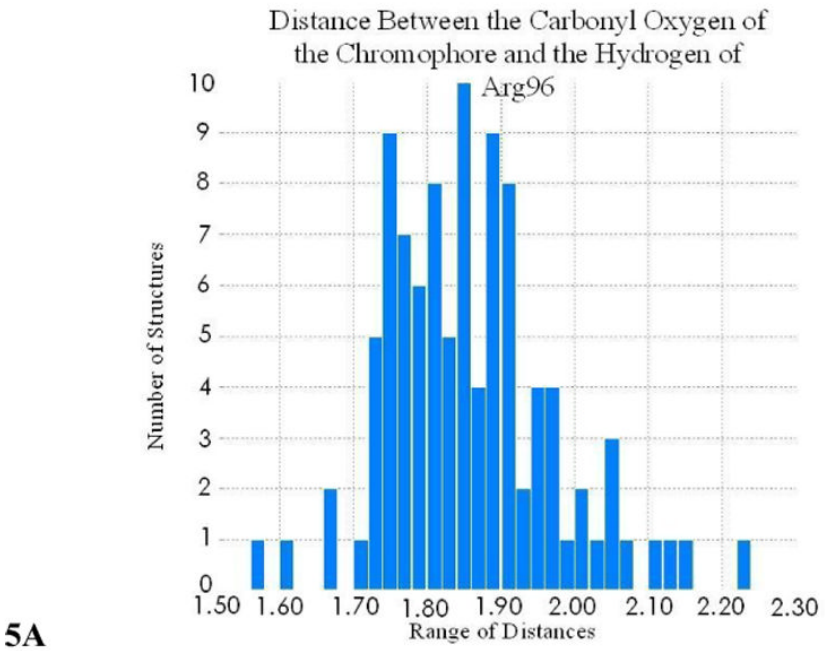
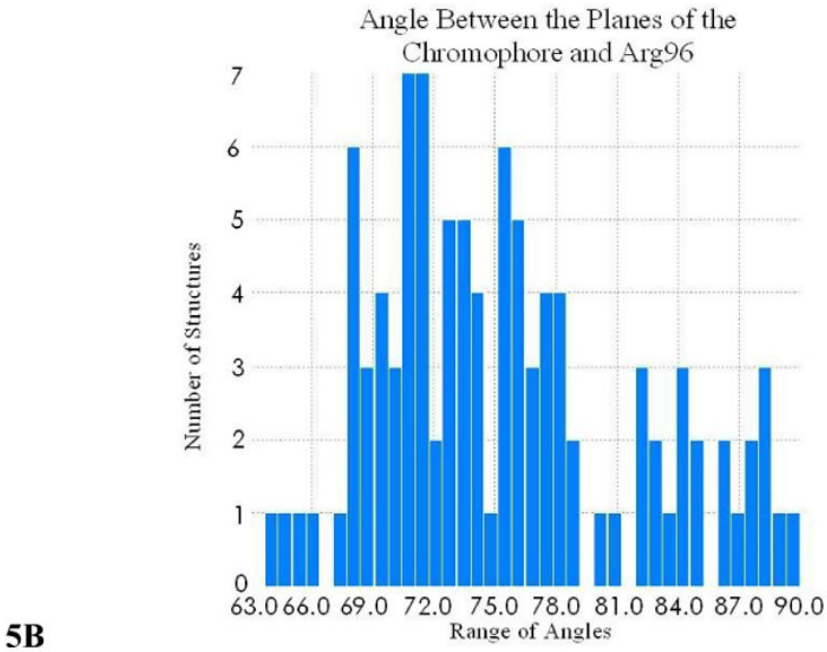


Figure 4.
An Isostar overlay plot of all the crystal structures of GFP and GFP-like proteins with fully cyclized chromophores found in the pdb (99 structures, see supplementary Table I for list of structures). The imidazolone rings of all the structures were overlapped in order to show the orientation of the arginine and glutamic acid residues (Arg96 and Glu222 in GFP numbering) relative to the position of the imidazolone ring.



5A



5B

Figure 5.
Distribution of distances (6A) between the carbonyl oxygen of Gly67 and Arg96 (GFP numbering), and the angles (6B) between the best plane of the guanidinium group of all the arginines and the imidazolone rings shown in Figure 5 for all 99 structures in the pdb with cyclized chromophores.

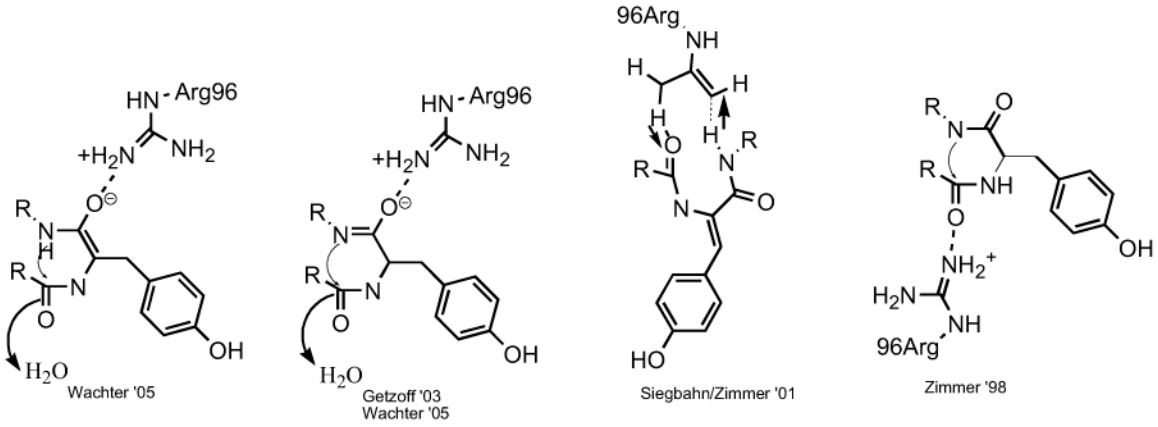


Figure 6.
Possible roles for Arg96 in the autocatalytic cyclization reaction responsible for chromophore formation.

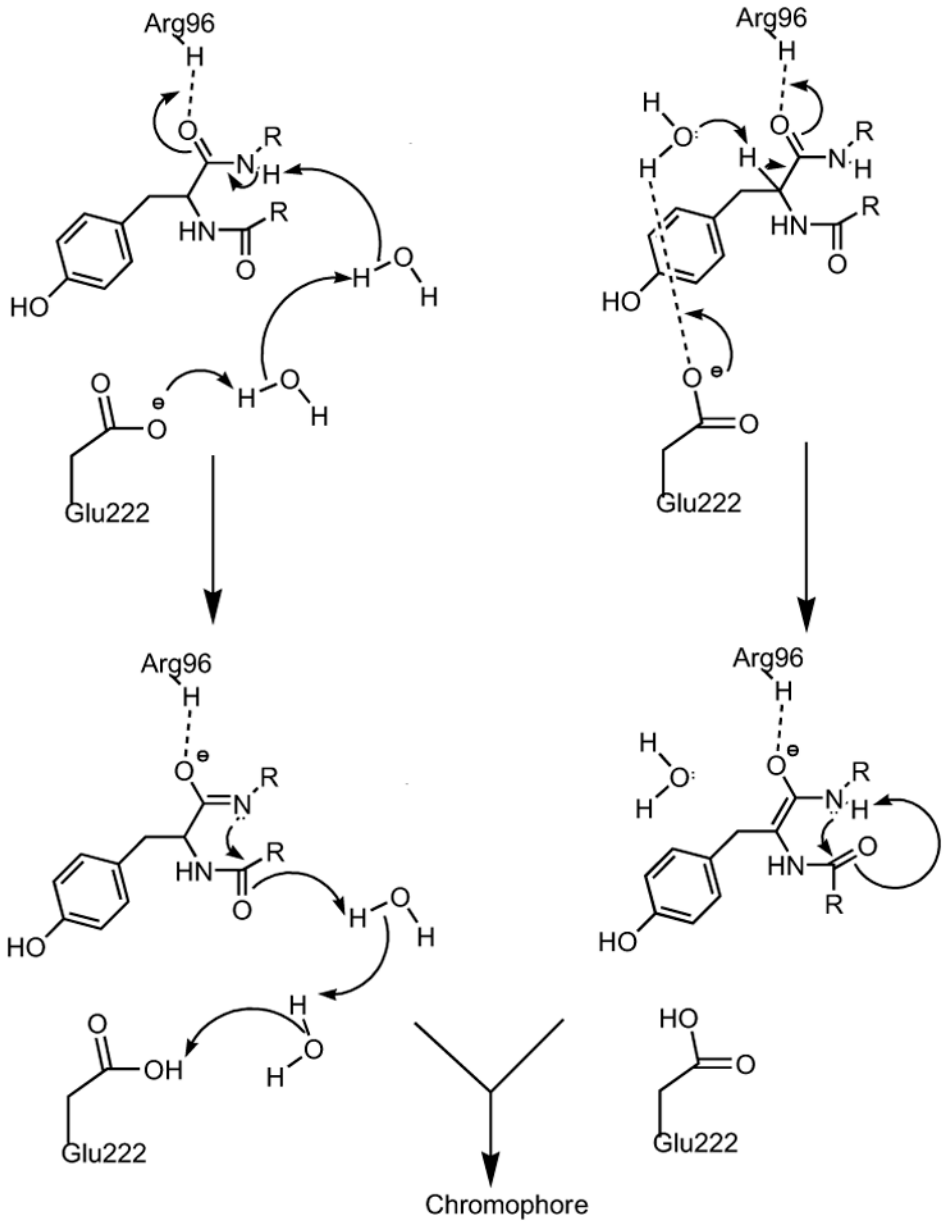


Figure 7.
Potential roles for Glu222. Backbone condensation may be initiated by deprotonation of the Gly67 nitrogen (left) or the Tyr66 α-carbon (right).

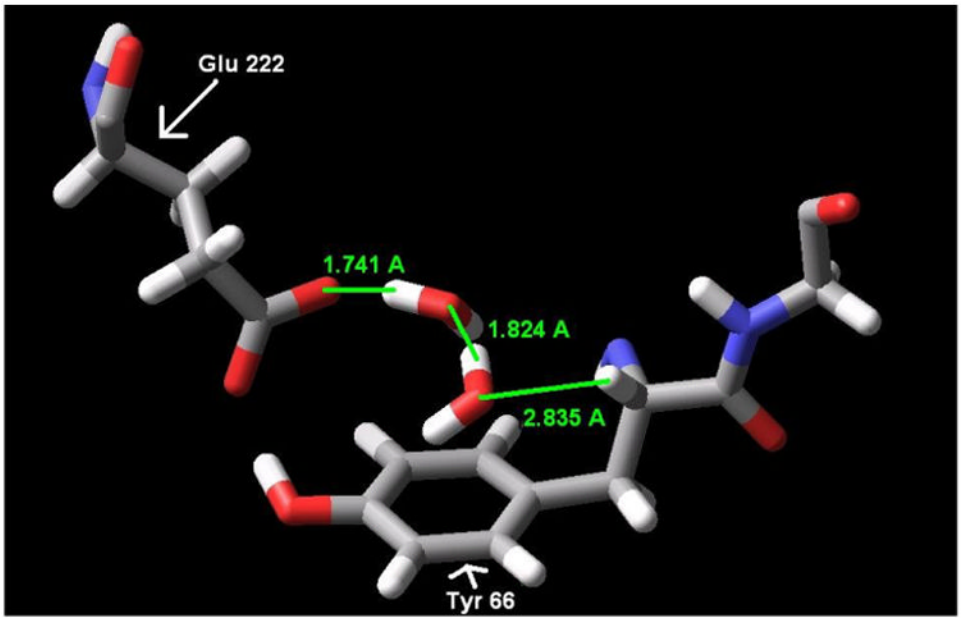


Figure 8.
Lowest energy conformation of immature GFP showing the proton shuttle from Tyr66 α-carbon to Glu222 showing that Glu222 can act as a general base, see Figure 7 right.

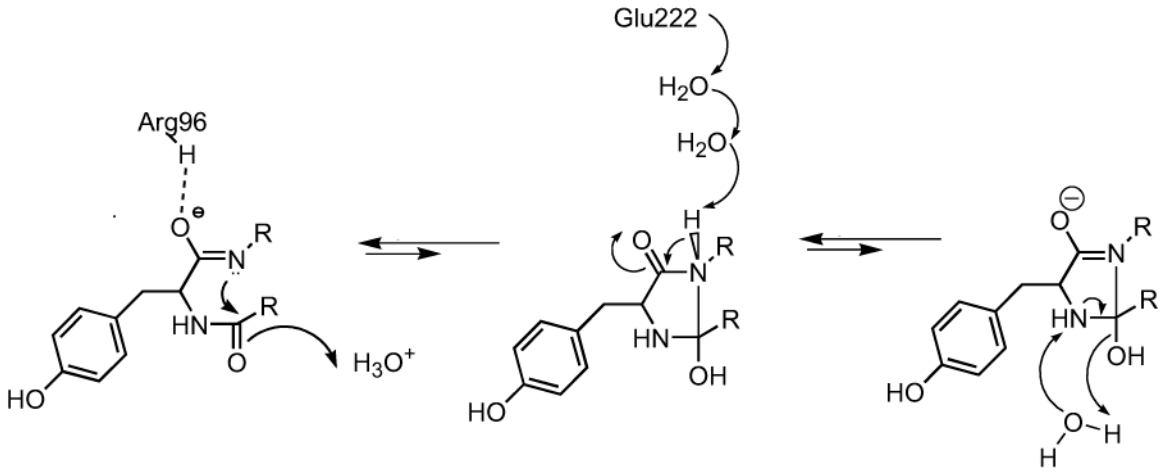


Figure 9.
The role of Glu222 in the first steps of the cyclization-dehydration-oxidation mechanism for chromophore biosynthesis.(14)

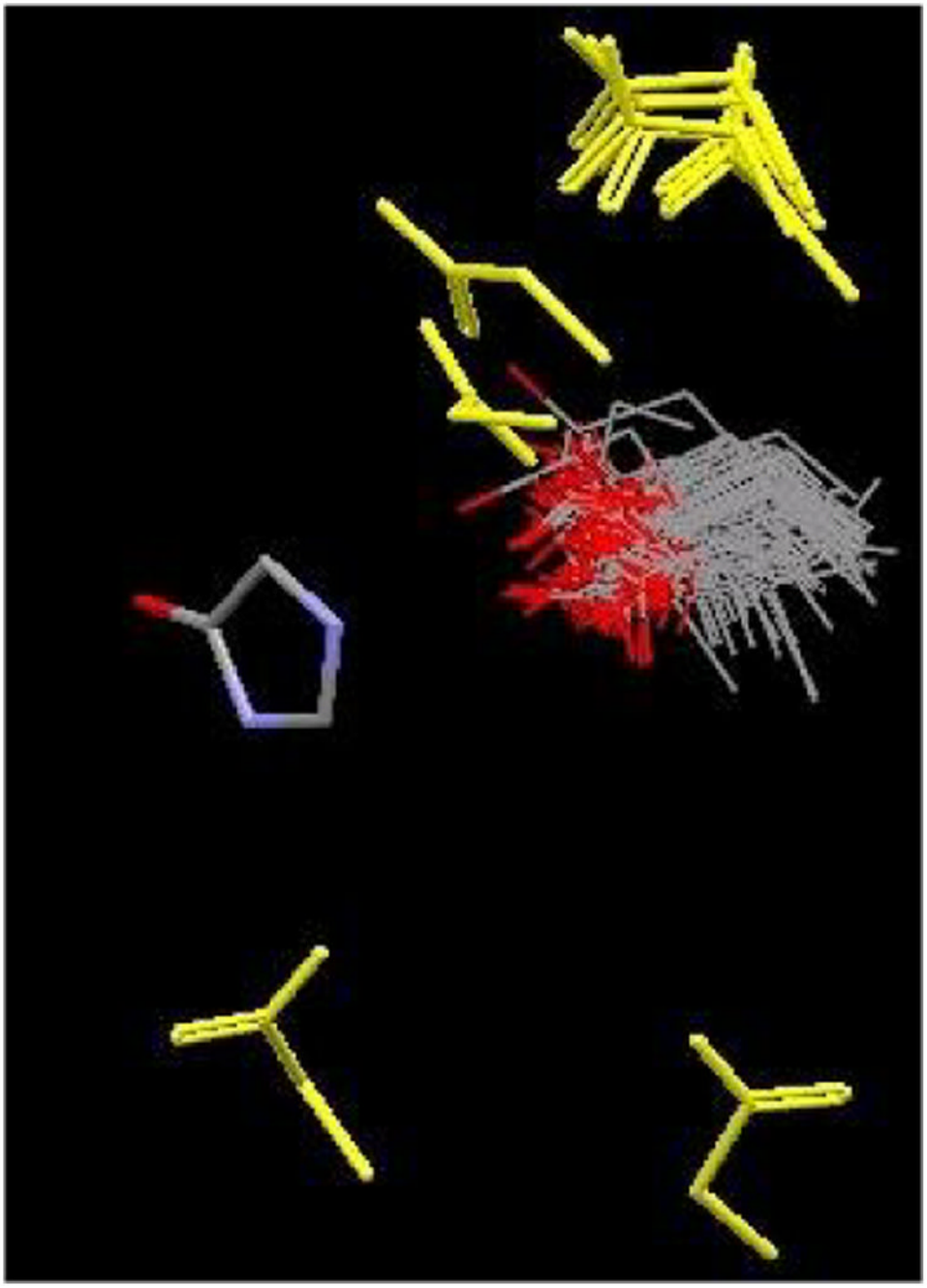
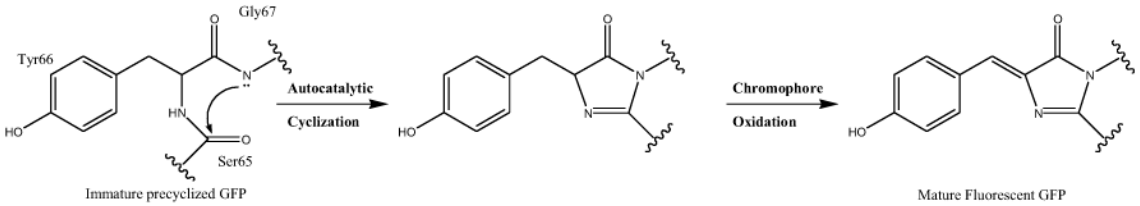


Figure 10.

An Isostar overlay plot of all the crystal structures of GFP and GFP-like proteins with fully cyclized chromophores found in the Protein Databank, as well as all the cyclized HAL, PAL and TAM structures. The imidazolone rings of all the structures were overlapped in order to show the orientation of the glutamic acid residues (Glu222 in GFP numbering) relative to the position of the imidazolone ring. The Glus closest to the imidazolone ring in all the HAL, PAL and TAM structures are colored yellow, and do not fall in the same space as the Glu222s in the chromophore-forming proteins (colored according to atom color).



Scheme 1.

Table I

Comparison of the geometric parameters associated with the tight turn motif in the 10,606 low energy structures and all the GFP structures with immature chromophores in the pdb, as well as all the interactions that would be hydrogen bonds in a canonical α-helix. Non-bonded interactions that meet the Maestro requirements set for hydrogen bonds are bold.

Structure	Tight Turn Distance (Å)	Distance (Å) between atoms (C=O …H-N) that would form a hydrogen bond in an ideal α helix							
		C_{60}-N_{64}	C_{61}-N_{65}	C_{62}-N_{66}	C_{63}-N_{67}	C_{64}-N_{68}	C_{65}-N_{69}	C_{66}-N_{70}	C_{67}-N_{71}
Average Calc.	3.115	3.045	2.592	4.126	7.255	3.726	4.868	5.620	5.931
Lowest energy	3.132	3.045	2.586	4.126	8.003	3.792	4.266	8.280	5.942
2AWJ – R96M	3.031	**1.848**	**2.186**	**2.117**	3.867	5.045	3.872	5.233	5.883
1QY3 – F64L, S65T, R96A	3.327	**1.923**	**2.147**	2.282	3.456	4.629	3.575	4.889	5.755
1YHG – S65G, Y66S, G67A	3.098	**1.762**	**1.878**	3.544	4.885	5.624	3.951	3.682	2.852
2HFC – F64L, S65T, Y66F, R96A	3.168	**1.919**	**2.085**	**2.182**	3.852	4.685	3.251	4.748	5.520
1YHI – S65A, Y66S, R96A	3.302	**2.248**	**1.951**	**2.383**	3.042	3.508	3.741	5.127	6.218
1QXT – F64L, S65T, R96A	3.031	**1.921**	**2.213**	**2.020**	4.022	4.556	3.558	5.147	5.778
1QYO – S65G, Y66G	3.019	**1.887**	**1.742**	2.858	3.341	4.343	4.931	6.584	6.844
1YHH - S65A, Y66S, G67A	3.038	**1.750**	**1.977**	3.353	4.660	5.910	4.049	3.883	2.811
1GFL - Mature	---	**1.986**	**1.977**	4.446	---	4.189	---	6.885	6.921

Table II

List of hydrogen bonds involving the main chain atoms of residues 64 – 68. The hydrogen bonds found in an ideal α-helix (see Table I) are not listed.

Structure	(C=O)64	(N-H)64	(C=O)65	(N-H)65	(C=O)66	(N-H)66	(C=O)67	(N-H)67	(C=O)68	(N-H)68
Lowest Energy	H_2O	Val61(C=O)	Val68(N-H) Gly67(N-H)		Gln94(N-H) Arg96(N-H)		H_2O Tyr92(OH) Gln69(N-H)		Phe71(N-H)	
2AWJ – R96M			Val68(N-H)	Val61(C=O)		Thr62(C=O)		H_2O	Phe71(N-H)	
1QY3 – F64L, S65T, R96A		Leu60(C=O)	Val68(N-H)	Leu60(C=O)	Gln69(N-H)				Phe71(N-H)	Thr65(C=O)
1YHG – S65G, Y66S, G67A			Gly68(N-H)	Val61(C=O)	Gln69(N-H)		Cys70(N-H)	H_2O	Phe71(N-H)	Gly65(C=O)
2HFC – F64L, S65T, Y66F, R96A		Leu60(C=O)	Val68(N-H)	Val61(C=O)	Gln69(N-H)	Thr62(C=O)			Phe71(N-H)	Thr65(C=O)
1YHI – S65A, Y66S, R96A		Leu60(C=O)	Val68(N-H)	Val61(C=O)		Thr62(C=O)	Asn121(N-H)		Phe71(N-H)	Ala65(C=O)
1QXT – F64L, S65T, R96A		Leu60(C=O)	Val68(N-H)	Val61(C=O)		Thr62(C=O)	Asn121(N-H)	H_2O	Phe71(N-H)	Thr65(C=O)
1QYO – S65G, Y66G		Leu60(C=O)	Val68(N-H)	Val61(C=O)	Arg96(N-H) Gln94(N-H)	Val61(C=O)			Phe71(N-H)	Gly65(C=O)
1YHH - S65A, Y66S, G67A		Leu60(C=O)	Val68(N-H)	Val61(C=O)	Gln69(N-H) Gln69(N-H Sidechain)	Thr63(C=O)	Cys70(N-H)		Phe71(N-H)	Ala65(C=O)
1GFL - Mature		Leu60(C=O)		Val61(C=O)	Arg96(N-H) Gln94(N-H)				Phe71(N-H)	H_2O

Table III

Distances between the carbonyl oxygen of the closest glutamic acid to the post-translationally cyclized tripeptide and the α-carbon of residue corresponding to Tyr66, and between the arginine closest to the post-translationally cyclized tripeptide and the carbonyl oxygen of the ring system.

Protein Name	Description	PDB Refcode	Ring Residues	Closest ARG	Distance to ARG (Å)	Closest GLU	Distance to GLU (Å)
Histidine Ammonia Lyase							
HAL	From *Pseudomonas putida*	1B8F[42]	142–144	109	16.587	414	5.785
HAL	Mutant Y280F from *Pseudomonas putida*	1GKJ[47]	142–144	109	16.620	414	6.322
HAL	From *Pseudomonas putida* Inhibited with L-Cysteine	1GKM[47]	142–144	109	16.864	414	5.275
HAL	Mutant F329A from *Pseudomonas putida*	1EB4[44]	142–144	109	16.624	414	6.151
Phenylalanine Ammonia Lyase							
PAL	From *Nostoc punctiforme*	2NYF[54]	167	128	8.726	448	9.101
PAL	From *Anabaena variabilis*	2NYN[54]	167	128	8.440	448	8.952
PAL	From *Rhodosporidium toruloides*	1T6J[52]	211	172	9.211	496	8.796
PAL	From *Rhodosporidium toruloides*	1T6P[52]	211	172	8.222	496	7.154
PAL	From *Petroselinum crispum*	1W27[55]	203	163	7.477	484	6.497
PAL	From yeast *Rhodosporidium toruloides*	1Y2M[56]	211–213	172	7.066	496	7.091
Tyrosine Aminomutase							
TAM	From *Streptomyces globisporus*	2OHY[57]	152	113	7.249	334	10.682

Chem Phys. Author manuscript; available in PMC 2008 December 9.