VOLUME 120, NUMBER 1
JANUARY 14, 1998

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

A Computational Analysis of the Unique Protein-Induced Tight Turn That Results in Posttranslational Chromophore Formation in Green Fluorescent Protein

Bruce R. Branchini, Amy R. Nemser, and Marc Zimmer*

Contribution from the Chemistry Department, Connecticut College, New London, Connecticut 06320

Received August 27, 1997

Abstract: A thorough conformational search of the chromophore-forming region of immature green fluorescent protein (GFP) revealed that it is preorganized in a unique conformation required for chromophore formation. This "tight turn" conformation has an i carbonyl carbon to $i + 2$ amide nitrogen distance of less than 2.90 Å with $\phi = 60 \pm 30°$ and $\psi = 30 \pm 15°$. Less than 1.00% of the residues of the 50 representative proteins examined adopt this conformation. The tight turn conformation is predominately located on the periphery of the proteins or in flexible areas, except in GFP. Molecular dynamics simulations and Ramachandran plots show the chromophore-forming region in immature GFP can only adopt the tight turn conformational family. Moreover, this conformation is ideally suited for the cyclization necessary for chromophore formation, i.e., for nucleophilic attack of the amino group of Gly67 on the carbonyl group of Ser67. The 11 β sheets of GFP force the chromophore-forming peptide fragment to adopt a conformation that has an exceptionally short interatomic distance between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67 and lock it into this conformation. Several mutant GFPs have been expressed that exhibit greater solubility and thermostability than wild-type GFP. These properties are linked to protein folding and chromophore formation. Our calculations show that the mutations cause a shortening of the distance between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67, and therefore enhance chromophore formation.

Introduction

The green bioluminescence of the jellyfish *Aequorea victoria* and sea pansy *Renilla reniformis* is due to the excitation of fluorescence from a chromophore found in the noncatalytic green fluorescent proteins (GFPs). The gene for green fluorescent protein from *Aequorea victoria* has been cloned,[1] and the expressed recombinant protein has been shown to be very similar to or identical with the native one.[2] GFP has been expressed in a wide variety of organisms such as *Escherichia coli*, *Caenor habditis elegans*,[3] *Xenopus laevu*, *Drosophilia melanogaster*,[4] zebra fish,[5] plants,[6] human embryonic kidney cells,[7] transgenic mice,[8] and other mammalian cells.[9] Currently

(1) Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Pendergast, F. G.; Cormier, M. J. *Gene* **1992**, *111*, 229.
(2) Inouye, S.; Tsuji, F. I. *FEBS Lett.* **1994**, *351*, 211.
(3) Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. *Science* **1994**, *263*, 802.
(4) Gerdes, H. H.; Kaether, C. *FEBS Lett.* **1996**, *389*, 44.
(5) Amsterdam, A.; Lin, S.; Hopkins, N. *Dev. Biol.* **1995**, *171*, 123.
(6) Chiu, W.-L.; Niwa, Y.; Zeng, W.; Hirano, T.; Kobayashi, H.; Sheen, *J. Curr. Biol.* **1996**, *6*, 325. Heimlein, M.; Epel, B. L.; Padgett, H. S.; Beachy, R. N. *Science* **1995**, *270*, 1983.
(7) Marshall, J.; Molloy, R.; Moss, G. W. J.; Howe, J. R.; Hughes, T. E. *Neuron* **1995**, *14*, 211.
(8) Chiocchetti, A.; Tolosano, E.; Hirsh, E.; Silengo, L.; Alturda, F. *Biochim. Biophys. Acta* **1997**, *1352*, 193.
(9) Ludin, B.; Doll, T.; Meill, R.; Kaech, S.; Matus, A. *Gene* **1996**, *173*, 113.

S0002-7863(97)03019-9 CCC: $15.00
Published on Web 01/14/1998

there is widespread interest in the application of GFP to *in situ* monitoring of gene expression, protein movement, and cell development.[10−13] GFP-tagged proteins can be monitored noninvasively in living cells by flow cytometry, fluorescence microscopy,[14] or macroscopic imaging methods. The advantages of GFP as a reporter protein are its natural fluorescence, stability, small size (238 amino acids (aa's)) which allows it to diffuse through extensively branched cells such as neurons,[15] and heterologous expression.[16] Additionally, GFP chromophore formation does not require any additional factors, the protein does not interfere with cell growth and function, and it is not toxic to cells. The disadvantages are that the onset of fluorescence after expression is slow and that there are two fluorescent excitation peaks.[3,17] These disadvantages have been partially overcome in the mutant GFP-S65T,[18] a commercially available reporter gene, and in other mutants,[19−22] making it a practical alternative to reporter genes such as *cat*, *lacZ*, and *luciferase*. Recently it has been shown that each individual GFP molecule can be switched optically between an emissive and nonemissive state.[23] As a consequence, GFP also might be used in the design of optical switching and optical storage devices, and perhaps to monitor time-dependent cell processes.

Proteolytic treatment of GFP demonstrated that the chromophore was contained in a stable hexapeptide fragment, 64FSYGVQ69.[24] In the intact GFP, the intrinsic chromophore is formed by autocatalytic internal cyclization of the tripeptide 65SYG67 and subsequent oxidation of the intrinsically formed structure. GFP fluorescence is not observed until 90 min to 4 h after protein synthesis.[25,30] Apparently, the protein folds quickly, but the subsequent fluorophore formation and oxidation are slow.[26] GFP refolding from an acid-, base-, or guanidine HCl-denatured state (chromophore-containing but nonfluorescent) occurs with a half-life of between 24 s[27] and 5 min,[28] and the recovered fluorescence is indistinguishable from that of native GFP.[29]

A most interesting feature of GFP is that its function is based on a chromophore formed through a rarely observed posttranslational cyclization of a peptide from its own backbone structure. The detailed mechanism for the formation of this structure is unknown. However, Tsien,[30,31] has proposed the autocatalytic

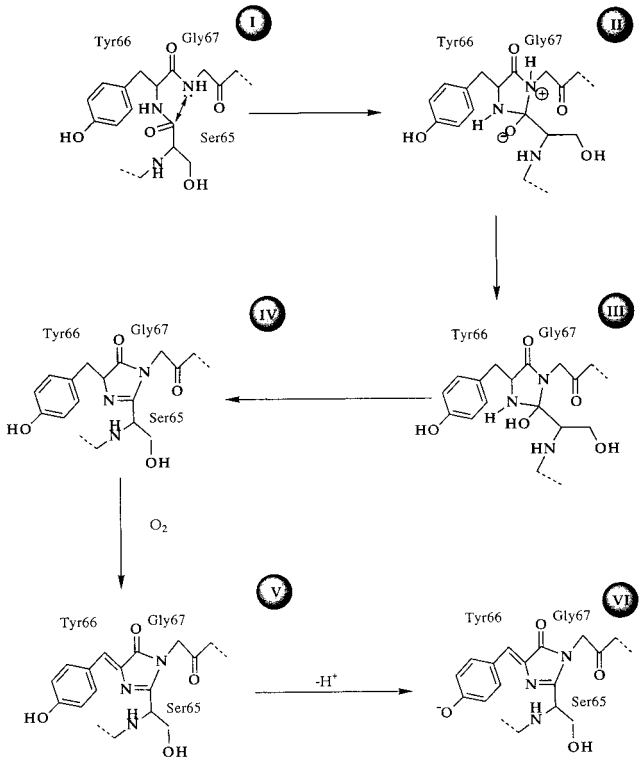

Figure 1. Proposed mechanism[30,31] for the chromophore formation in green fluorescent protein.

biosynthetic mechanism shown in Figure 1. The scheme accounts for the spontaneous chromophore formation in a variety of GFP-expressing organisms which are unlikely to contain the same specific catalysts for the process. In this paper we will refer to fluorescent chromophore-containing GFP as mature GFP and to unmodified primary structure as immature GFP.

Recently the crystal structures of wild-type GFP as both a dimer[32] and a monomer,[33] and the solid-state structures of several mutants,[34,35] have been reported. The structure of GFP has been described as a *light in a can*; the chromophore is located in the center of a can consisting of 11 β sheets. The can is a nearly perfect cylinder with a height of 42 Å and a radius of 12 Å; see Figure 2. By being enclosed in the can, the chromophore may be protected from quenching by oxygen[36] and attack by hydronium ions.[28] Deletion mapping experiments[37] have shown that nearly the entire structure (residues 2−232) is required for chromophore formation and/or fluorescence.

On the basis of the computational analyses of the hexapeptide FSYGVQ, which were completed prior to the publication of the crystal structure of GFP, we proposed[38,39] that the posttranslational chromophore formation occurs due to the presence of low-energy conformations which have very short intramolecular distances between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67 (↔ in **I**, Figure 1). We also suggested

(10) Santa Cruz, S.; Chapman, S.; Roberts, A. G.; Roberts, I. M.; Prior, D. A. M.; Oparka, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6286.
(11) Gerdes, H.-H.; Kaether, C. *FEBS Lett.* **1996**, *389*, 44.
(12) Tannahill, D.; Bray, S.; Harris, W. A. *Dev. Biol.* **1995**, *168*, 12501.
(13) Wang, S.; Hazelrigg, T. *Nature* **1994**, *369*, 400.
(14) Gilroy, S. *Annu. Rev. Plant Physiol. Mol. Biol.* **1997**, *48*, 165.
(15) Brand, A. *Trends Genet.* **1995**, *11*, 324.
(16) Gura, T. *Science* **1997**, *276*, 1989.
(17) Chalfie M. *Photochem. Photobiol.* **1995**, *62*, 651.
(18) Heim, R.; Cubitt, A. B.; Tsien, R. Y. *Nature* **1995**, *373*, 663.
(19) Cormack, B. P.; Valdivia, R. H., Falkow, S. *Gene* **1996**, *173*, 33.
(20) Reichel, C.; Mathur, J.; Eckes, P.; Langenkemper, K.; Koncz, C.; Schell, J.; Reiss, B.; Maas, C. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5888.
(21) Yang, T.-T.; Cheng, L.; Kain, S. R. *Nucleic Acids Res.* **1996**, *24*, 4592.
(22) Crameri, A.; Whitehorn, E.; Tate, E.; Stemmer, W. *Nat. Biotechnol.* **1996**, *14*, 315.
(23) Dickson, R. M.; Cubitt, A. B.; Tsien, R. Y.; Moerner, W. E. *Nature* **1997**, *388*, 355.
(24) Cody, C. W.; Prasher, D. C.; Westler, W. M.; Pendergast, F. G.; Ward, W. W. *Biochemistry* **1993**, *32*, 1212.
(25) Crameri, A.; Whitehorn, E. A., Tate, E.; Stemmer, W. P. C. *Nat. Biotechnol.* **1996**, *14*, 315−319.
(26) Reid, B. G.; Flynn, G. C. *Biochemistry* **1997**, *36*, 6786.
(27) Makino, Y.; Amada, K.; Taguchi, H.; Yoshida, M. *J. Biol. Chem.* **1997**, *272*, 12468.
(28) Ward, W. W.; Bokman, S. H. *Biochemistry* **1982**, *21*, 1, 4535.
(29) Bokman, S. H.; Ward, W. W *Biochem. Biophys. Res. Commun.* **1981**, *101*, 1372.
(30) Heim, R.; Prasher, D. C.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12501.
(31) Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y. *Trends Biochem. Sci.* **1995**, *20*, 448.
(32) Yang, F.; Moss, L. G.; Phillips, G. N., Jr. *Nat. Biotechnol.* **1996**, *14*, 1246.
(33) Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormoe, M.; Remington, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2306.
(34) Ormoe, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. *Science* **1996**, *273*, 1392.
(35) Palm, G.; Zdanov, A.; Gaitanaris, G. A.; Stauber, R.; Pavlakis, G. N.; Wlodawer, A. *Nat. Struct. Biol.* **1997**, *4*, 361.
(36) Nageswara Rao, B. D.; Kempel, M. D.; Pendergast, F. G. *Biophys. J.* **1980**, *32*, 630.
(37) Dopf, J.; Horiagon, T. M. *Gene* **1996**, *173*, 39.

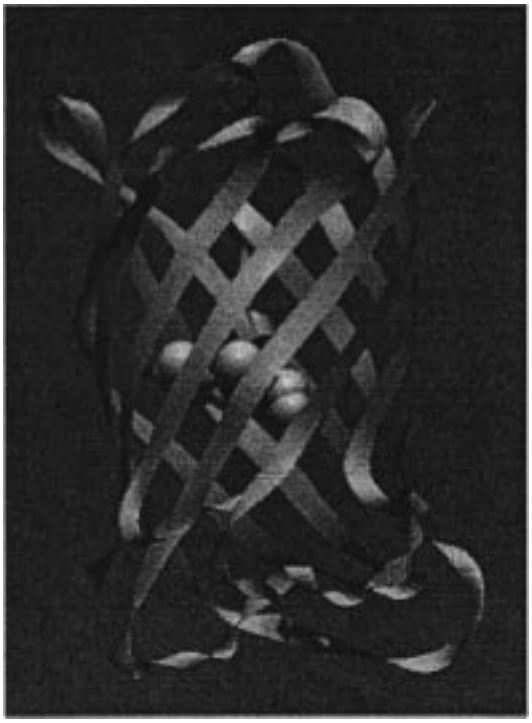

Figure 2. Solid-state structure of GFP. The chromophore is located in the center of the can and is shown with a CPK representation. (Coordinates for the figure were obtained from the PDB, code 1GFL.)

that an arginine side chain may hydrogen bond to the carbonyl oxygen of Ser65, activating the carbonyl carbon of Ser65 for attack by the lone pair of the Gly67 amide nitrogen.[38,39] Theclose proximity of an arginine, namely, Arg96, has since been confirmed by the GFP crystal structures.

All of the solid-state structures of GFP[32−35] show that it is compromised of an 11 β sheet barrel surrounding the chromophore as displayed in Figure 2. The crystal structures of GFP and its mutants have been used to show that the two absorption maxima in GFP correspond to the neutral and anionic chromophore states (Figure 1, **V** and **VI**).[33] Although the available structures do not shed much light on the mechanism of the chromophore formation, they provide useful starting points for the computational analysis of the autocatalytic cyclization shown in Figure 1.

In this paper, we use the solid-state structure[32] of wild-type GFP to identify key GFP structural features which we propose are the basis for chromophore formation, namely, that the 11 β sheets force the chromophore-forming peptide fragment to adopt a conformation that has an exceptionally short interatomic distance between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67 and lock it into this conformation.

Experimental Section

The coordinates of the wild-type GFP solid-state structure (1GFL)[32] were obtained from the Protein Data Bank;[40] hydrogen atoms were added to protein and solvent atoms as required. The chromophore was graphically restored to the polypeptide as it would be prior to the autocatalytic cyclization (i.e., from species **IV** to species **I** in Figure 1 or Figure 3b to Figure 3a). Unless otherwise specified, crystallographically determined water molecules were incorporated in the calculations. A "hot" sphere of 12 Å from residues 65−67, with a secondary constrained sphere extending a further 3 Å, was used in all simulations and conformational searches. Constraints were introduced with a harmonic restoring potential of 100 kJ/Å.^{2} The AMBER* force field as implemented in MacroModel v5.5[41] was used for all molecular modeling. It uses a 6−12 Lennard-Jones hydrogen-bonding treatment[42] and an improved protein backbone parameter set.[43]

Dihedral Monte Carlo multiple minimum searches[44] of the chromophore-forming region in GFP were undertaken with the closure bonds and the rotatable dihedral angles shown in Figure 4. A minimum ring closure distance of 1.00 Å and a maximum of 4.00 Å were used. Each Monte Carlo step varied between 1 and 15 of the rotatable dihedral angles by between 0 and 180°. During the search procedure minimization continued until convergence was reached or until 1000 iterations

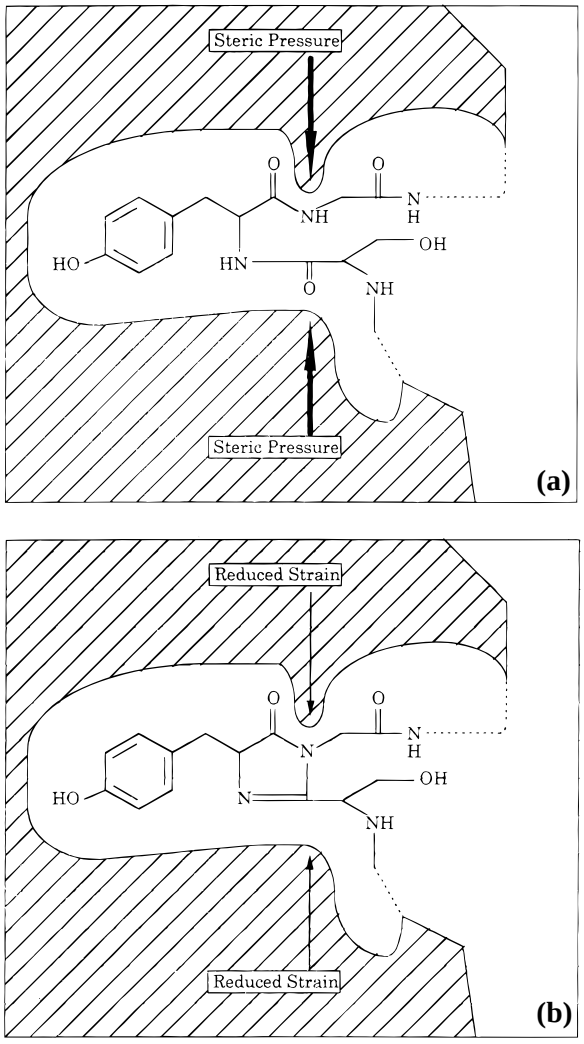

Figure 3. (a, top) All calculations were done on the immature precatalytic green fluorescent protein represented in this figure. The arrows depict the strain exerted by the remainder of the protein, which enforces the tight turn conformation and thereby preorganizes the chromophore-forming residues for the autocatalytic ring closure. (b, bottom) The crystal structures of GFP are all of the mature form of GFP which is depicted in this figure and is formed by a posttranslational modification of GFP.

(38) Zimmer, M., Branchini, B., Lusins, J. O. In *Bioluminescence and Chemiluminescence, Proc. 9th Int. Symp.*; Hastings, J. W., Kricka, L. J., Stanley, P. E., Eds.; Wiley: New York, 1990; p 407.
(39) Branchini, B. R., Lusins, J. O., Zimmer, M. *J. Biomol. Struct. Dyn.* **1997**, *14*, 441.
(40) Bernstein, F. C.; Koetzle, T. F.; Williams, J. G. B.; Meyer, E. F.; Brice, M. R.; Rodgers, J. R.; Kennard, O.; Shimanouchi T.; Tasumi, M. *J. Mol. Biol.* **1977**, *122*, 535.
(41) Mohamadi, F.; Richards, N. G. F.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caulfield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.
(42) Ferguson, D. M.; Kollman, P. A. *J. Comput. Chem.* **1991**, *12*, 620.
(43) McDonald, Q.; Still, W. C. *Tetrahedron Lett.* **1992**, *33*, 7743.
(44) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *11*, 4379.

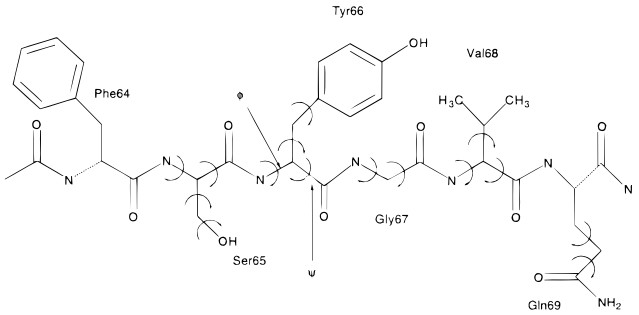

Figure 4. All the dihedral angles rotated in the dihedral Monte Carlo multiple minimum searches (arrows), the two closure bonds (⋯), and the dihedral angles ϕ and ψ that were varied in the Ramachandran plots.

had been performed. The Polak−Ribiere conjugate gradient minimization mode was used "in vacuo" with a derivative convergence criterion of 0.05 kJ/mol. Structures within 50 kJ/mol of the lowest energy minimum were kept, and a usage-directed method[45] was used to select structures for subsequent MC steps. Six thousand Monte Carlo steps were taken, and all conformations within 50 kJ/mol of the lowest energy conformation were combined, and subjected to a further 10 000 iterations with the multiconformer minimization mode of MacroModel. All unique conformations within 50 kJ/mol of the global minimum structure were kept. Structures were considered unique when the least squares superimposition of all the pairs of related non-hydrogen atoms in residues 64−69 found no pair that was separated by less than 0.025 Å. Using MacroModel v6.0, Ramachandran plots were generated by driving the ϕ and ψ dihedral angles shown in Figure 4 with 1000 kJ/mol torsional constraints.

Conformational searches were also carried out by sampling structures during molecular dynamic simulations. The simulations were run for 1.0 ns at 750 K using the SHAKE algorithm[46,47] with 1.5 fs time steps. Structures were sampled every 1.0 ps, resulting in 1000 structures which were minimized with the multiconformer minimization mode of MacroModel. All unique conformations within 50 kJ/mol of the global minimum structure were kept. Structures were considered unique when the least squares superimposition of all the pairs of related non-hydrogen atoms in residues 64−69 found no pair that was separated by less than 0.025 Å. The molecular dynamics and the dihedral Monte Carlo multiple minimum conformational searches used the "hot" and constrained zones described above.

Conformational families of the chromophore-forming region within GFP were established with the agglomerative, hierarchical single-link clustering program Xcluster.[48] Proximity matrixes were obtained by determining the pairwise distances between heteroatoms in residues 64−69 after optimal rigid-body superimposition.[49]

The distances from the carbonyl carbon of residue *i* to the amide nitrogen of residue $i + 2$ were determined with a script (see the Supporting Information) for the VMD program.[50] A nonhomologous set of 50 proteins was used in the analysis (see the Supporting Information). The proteins were selected from the work of Hutchinson[51] so that no two chains had more than 35% sequence identity. Structural analogues were removed using the structural alignment program SSAP.[52]

Small-molecule searches were performed using the Cambridge Structural Database (CSD) version 5.11 which contains 152 464 structures and was last updated in April 1996.

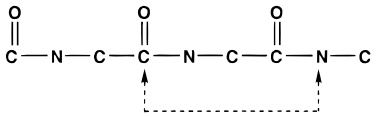

Figure 5. Fragment used in the Cambridge Structural Database search. The interatomic distance between the *i* carbonyl carbon and the $i + 2$ nitrogen was recorded for each hit.

Results and Discussion

Determination of the Low-Energy Conformations of the Chromophore-Forming Region within Immature Green Fluorescent Protein. The chromophore in the solid-state structure of wild-type GFP[32] was computationally converted from the mature GFP solid-state structure to immature precatalytic GFP (i.e., changed from species **V** to species I in Figure 1 and Figure 3b to Figure 3a). A thorough dihedral Monte Carlo multiple minimum search was undertaken to find the most stable conformations of GFP prior to chromophore formation. Six thousand Monte Carlo steps were taken, and the resulting conformations were minimized and combined as described in the Experimental Section. One thousand sixty-nine unique conformations were obtained. The lowest energy conformation had a distance of 2.873 Å between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67. A similar short distance was found in all low-energy conformations. In fact, all 500 lowest energy conformations had distances within 0.01 Å of 2.873 Å.

The Low-Energy Conformations of the Chromophore-Forming Region within Immature Green Fluorescent Protein Are All in a Similar Unique Conformation. To show that the *N*H to *C*O distances are unusually short and are of significance, we searched the Cambridge Structural Database (CSD) and Protein Databank (PDB) for their occurrence in other peptides. The search of the Cambridge Structural Database, which contains all the published crystal structures of molecules with 500 and fewer atoms, including many cyclic peptides, revealed that only 1.38% of molecules containing the fragment shown in Figure 5 have an interatomic distance between the carbonyl carbon of residue *i* and the nitrogen of residue $i + 2$ of less than 3.00 Å. Moreover, none of the 2460 structures have a distance shorter than 2.90 Å. A similar search of 50 representative proteins selected from the PDB showed that only 0.89% of 18 107 *i* carbonyl carbon to $i + 2$ nitrogen distances were less than 2.90 Å (see the Supporting Information). We have called this motif in which there is an *i* carbonyl carbon to $i + 2$ nitrogen distance of less than 2.90 Å a "tight turn" conformation. It is different from any of the reported reverse turns.[53] Most tight turns appear on the surface of the protein, and those that are found in the interior are also flexible and are not locked into this tight turn conformation.

The low-energy conformations of the FSYGVQ hexapeptide fragment in solution were compared with the low-energy conformations of the same hexapeptide sequence located in GFP. The lowest energy tight turn conformation of FSYGVQ in solution was 5.87 kJ/mol higher in energy than the global energy minimum.[39] Furthermore, no conformations with distances of less than 2.90 Å were found for the hexapeptide fragment; the shortest distance found was 2.94 Å. *However, we will show (vide infra) that all the low-energy conformations of FSYGVQ in GFP are in the tight turn conformation. These observations*

(45) Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419.
(46) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(47) Ryckaert, J. P. *Mol. Phys.* **1985**, *55*, 549.
(48) Shenkin, P. S.; McDonald, D. Q. *J. Comput. Chem.* **1994**, *15*, 899.
(49) Kabsh, W. *Acta Crystallogr.* **1976**, *A32*, 922. Kabsh, W. *Acta Crystallogr.* **1978**, *A32*, 827.
(50) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33. The program is available from http://www.ks.uiuc.edu/Research/vmd.
(51) Hutchinson, E. G.; Thornton, J. M. *Protein Sci.* **1994**, *3*, 2207.
(52) Orengo, C. A.; Brown, N. P.; Taylor, W. R. *Proteins: Struct., Funct., Genet.* **1992**, *14*, 367.
(53) Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2nd ed.; W. H. Freeman: New York, 1993.

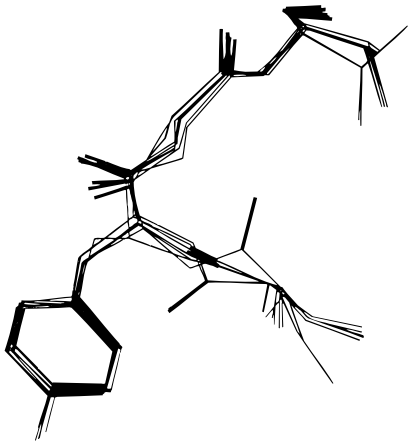

Figure 6. Representative structures from each of the eight conformational families overlapped to show that they are all fairly similar and are members of the tight turn family of conformations.

indicate that the GFP protein enforces the tight turn conformation, and therefore preorganizes the chromophore-forming residues for the autocatalytic ring closure as shown in Figures 1 and 3a.

Cluster analysis was used to group the 1069 conformations obtained from the dihedral Monte Carlo multiple minimum search of the chromophore-forming region of immature GFP into conformational families. The aim of a cluster analysis is to place objects into groups, also called clusters, in such a way that all the objects within a cluster are very similar and that all the objects in different clusters are very dissimilar to each other. In this study we have used the xcluster program[48] which is an agglomerative, hierarchical, single-link method. The separation ratio,[54] distance maps, and mosaics were used to establish that there are eight distinct conformations (clusters) that differ more significantly from each other than any other conformations. The largest conformational family for immature GFP contains all the lowest energy conformations, and the average distance between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67 is 2.873 Å. In fact all the members of this group have distances within 0.01 Å of 2.873 Å. The remaining conformational families are less populated and have slightly longer distances. The largest of these families has 168 members and a distance of 2.965 Å and is 2.10 kJ/mol higher in energy than the average lowest energy conformation. An overlap of residues 64−69 of representative structures from the 8 families, Figure 6, shows that the conformations are fairly similar.

The database searches have shown that the tight-turn conformation in GFP, which a cluster analysis of all the low-energy conformations has shown to be the only conformation available to the chromophore-forming region of GFP, is very rare in other proteins.

The Chromophore-Forming Region within Immature Green Fluorescent Protein Is Preorganized in a Rigid Unique Tight Turn Structure. While the distance between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67 in the 500 lowest energy conformations in GFP varies no more than 0.01 Å, the same interatomic distance in the lowest 500 energy conformations of the free hexapeptide varies by more than 1.90 Å. The protein therefore limits the range of conformations available to the residues in the chromophore-forming region.

The same conclusion can be drawn by comparing the

(54) Minimum separation ratio 2.12.

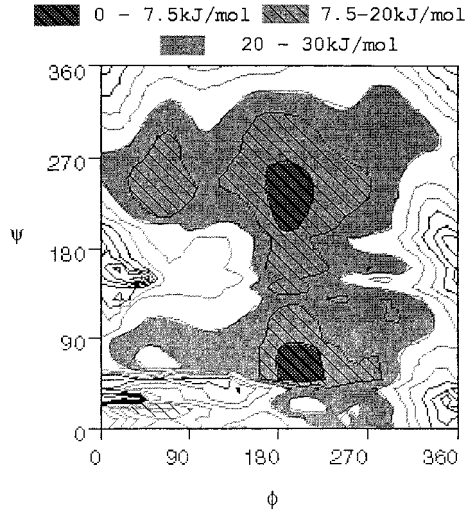

Figure 7. A Ramachandran plot generated by driving the ϕ and ψ dihedral angles of the tyrosine in the free hexapeptide FSYGVQ through 360° in 10° increments in a continuum of water.

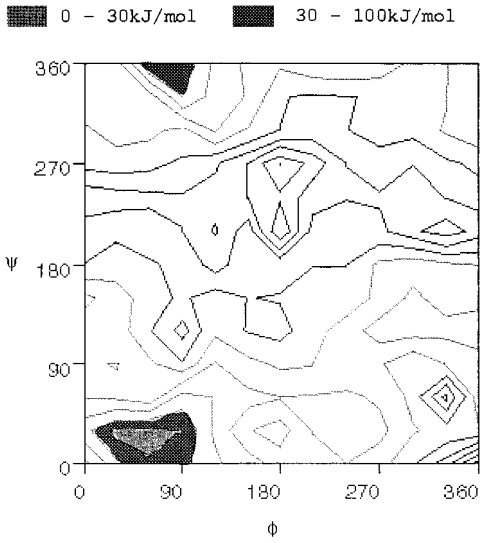

Figure 8. A Ramachandran plot generated by driving the ϕ and ψ dihedral angles of Tyr66 through 360° in 30° increments using the same "hot" and constrained zones of GFP described in the Experimental Section.

Ramachandran plots[55] shown in Figures 7 and 8. These plots were generated by rotating the ϕ and ψ dihedral angles shown in Figure 4. The Tyr66 ϕ and ψ dihedral angles were chosen as these are the torsions that are the best indicators of the tight turn conformation. The Ramachandran plot for the FSYGVQ peptide in a continuum of water, Figure 7, shows that the hexapeptide can adopt many different conformations; in fact more than half the ϕ/ψ space is within 20 kJ/mol of the lowest energy conformation. The Ramachandran plot for the same FSYGVQ sequence located within the β barrel of GFP, Figure 8, shows that the ϕ/ψ space is extremely restricted. The peptide can only adopt conformations with $\phi = 60 \pm 30°$ and $\psi = 30 \pm 15°$; these are the tight turn conformations. There are no other conformations within 100 kJ/mol of the tight turn conformation.

While the free hexapeptide can adopt a low-energy tight turn conformation, this conformation is not the one lowest in energy. Molecular dynamics simulations and Ramachandran plots reveal

(55) Ramachandran, G. N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283.

that the hexapeptide fragment is flexible and can easily interconvert among many different conformations, not staying in the tight turn conformation for a significant period of time. On the other hand, in immature GFP, the same six amino acid residues can **only** adopt the tight turn conformational family. Moreover, the chromophore-forming region of GFP is locked in a conformation that is ideally suited for nucleophilic attack of the amino group of Gly67 on the carbonyl group of Ser67.

To confirm that the dihedral Monte Carlo multiple minimum search found all the low-energy conformations, conformational searches also were carried out by sampling structures during molecular dynamic simulations. The simulations were run for 1.0 ns at 750 K, and the structures were sampled and minimized. All the conformations determined in this manner were identical with the conformations found in cluster 1 described above. The molecular dynamics search was not able to find the higher energy conformations located in clusters 2−8 and obtained in the dihedral Monte Carlo multiple minimum searches.

Confirmation of the Computational Results by Comparison with Empirical Data. All of the reported mutations at Gly67 have produced nonfluorescent proteins.[19] To establish if the glycine residue is crucial to the adoption of the tight turn conformation and therefore to chromophore formation, we replaced Gly67 with all the naturally occurring amino acids, and measured the distance between the carbonyl carbon of Ser65 and the amide nitrogen of aa67. This distance is significantly shorter in the wild type than in all the other distances that were obtained from the structures generated by mutating residue 67, doing a conformational search, and minimizing. The shortest carbonyl carbon to amide nitrogen distance was obtained for the G67A mutant, which is 0.05 Å longer than that observed in the wild type.

Additional experimental findings support our findings that chromophore formation in GFP occurs due to a precatalytic conformation that has a very close contact between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67, and because the conformational space of the protein in this region is severely restricted. During protein expression improper protein folding can result in aggregation and subsequent formation of insoluble inclusion bodies.[56] When GFP is expressed at 25 °C, most of the protein is soluble. Separation of the soluble and the insoluble fractions reveals that the soluble fraction fluoresces while the much smaller insoluble fraction does not. Expression of GFP in bacteria,[57] yeast,[58] and mammalian cells[59] is reduced at incubation temperatures greater than 30 °C because the amount of the insoluble nonfluorescing fraction has increased.[60] The fluorescence of mature GFP is temperature insensitive. In fact, mature GFP is a highly stable molecule whose in vitro fluorescence is unaffected by temperatures up to 60 °C.[28] Siemering has recently suggested that the temperature sensitivity of newly expressed GFP is due to the failure of immature GFP to fold into its autocatalytic conformation at higher temperatures.[60] Two single-point mutations, V163A and F64L, have been found that lead to higher yields of soluble fluorescent protein[35] and improve the levels of fluorescence in mammalian cells grown at 37 °C. If we assume that wild-type GFP chromophore formation at 37 °C is inefficient because the carbonyl carbon of Ser65 is not held close enough to the amide nitrogen of Gly67 for a sufficient length of time for nucleophilic attack to occur, and that this is due to increased movement and an increase in the population of higher energy conformations at higher temperatures, then the V163A and F64L mutants should affect the tight turn conformation by counteracting these effects. Using the same methods we used for immature native GFP, we found the low-energy conformations for the chromophore-forming regions of the immature V163A and F64L mutants. The distance between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67 in these mutants is significantly shorter than in the native GFP; see Table 1. This result may explain how the V163A and F64L mutations result in an increased yield of soluble fluorescent protein. In turn, these results provide further confirmation that a conformation that is held in place rigidly with a short carbonyl carbon of Ser65 to amide nitrogen of Gly67 distance is required for chromophore formation.

Table 1. Distances between the Carbonyl Carbon of Ser65 and the Amide Nitrogen of Gly67 in Immature Native GFP and in the V163A and F64L Mutants

species	carbonyl carbon of Ser65 to amide nitrogen of Gly67 distance (Å)
GFP (with water)	2.873
V163A	2.792
F64L	2.817

Conclusion

We have shown that the chromophore-forming residues of GFP are preorganized in a tight turn with the carbonyl carbon of Ser65 2.87 Å from the amide nitrogen of Gly67. Not only does the protein enforce this tight turn that is required in chromophore formation, it also tremendously restricts the conformational space of the chromophore-forming region, so that the residues are kept in place for autocatalytic cyclization, a slow step ($t_{1/2} \approx 5$ min) in chromophore formation. Several mutants have been expressed that exhibit greater solubility and thermostability, two properties that are linked to protein folding and chromophore formation. Our calculations show that the mutations decrease the carbonyl carbon of Ser65 to amide nitrogen of Gly67 distance and therefore enhance chromophore folding.

Acknowledgment. We thank Andrew Dalke for writing the VMD script and Masayaki Takahashi for help with the figures. M.Z. acknowledges support from the PRF (ACS-PRF Grant 31562-B3).

Supporting Information Available: Table listing the PDB codes, number of hits, and percent of hits with interatomic distance between the *i* carbonyl carbon and the $i + 2$ nitrogen of less than 2.90 Å and script for the VMD program used to determine the distances from the carbonyl carbon of residue *i* to the amide nitrogen of residue $i + 2$ in protein structures obtained from the PDB (4 pages). See any current masthead page for ordering and Internet access instructions.

JA973019J

(56) Kane, J. F.; Hartley, D. L. *Trends Biotechnol.* **1988**, *6*, 95.
(57) Webb, C. D.; Decatur, A.; Teleman, A.; Losick, R. *J. Bacteriol.* **1995**, *177*, 5906.
(58) Lim, C. R.; Kimata, Y.; Oka, M.; Nomaguchi, K.; Kohno, K. *J. Biochem.* **1995**, *118*, 13.
(59) Ogawa, H.; Inouye, S.; Tsuji, F.; Yasuda, K.; Umesono, K. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11899.
(60) Siemering, K. R.; Golbik, R.; Sever, R.; Haseloff, J. *Curr. Biol.* **1996**, *6*, 1653.