Chemical Physics 429 (2014) 5–11

Contents lists available at ScienceDirect

Chemical Physics

journal homepage: www.elsevier.com/locate/chemphys

Structural consequences of chromophore formation and exploration of conserved lid residues amongst naturally occurring fluorescent proteins

Matthew H. Zimmer [a,b], Binsen Li [a], Ramza Shahid [a], Paola Peshkepija [a], Marc Zimmer [a,*]

[a] Chemistry Department, Connecticut College, New London, CT 06320, USA
[b] Department of Chemistry, University of Oxford, Oxford OX1 3QZ, United Kingdom

ARTICLE INFO

Article history:
Received 8 October 2013
In final form 20 November 2013
Available online 6 December 2013

Keywords:
GFP
Aequorea
Folding

ABSTRACT

Computational methods were used to generate the lowest energy conformations of the immature precyclized forms of the 28 naturally occurring GFP-like proteins deposited in the pdb. In all 28 GFP-like proteins, the beta-barrel contracts upon chromophore formation and becomes more rigid. Our prior analysis of over 260 distinct naturally occurring GFP-like proteins revealed that most of the conserved residues are located in the top and bottom of the barrel in the turns between the β-sheets (Ong et al. 2011) [1]. Structural analyses, molecular dynamics simulations and the Anisotropic Network Model were used to explore the role of these conserved lid residues as possible folding nuclei. Our results are internally consistent and show that the conserved residues in the top and bottom lids undergo relatively less translational movement than other lid residues, and a number of these residues may play an important role as hinges or folding nuclei in the fluorescent proteins.

1. Introduction

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

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

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

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

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

GFP-like proteins fold into a distinctive beta barrel shape composed of 11 β-sheets surrounding a central alpha helix, which contains the chromophore, see Figs. 1 and 2. In *Aequorea victoria* GFP (avGFP), the chromophore is formed by an autocatalytic cyclization of the tripeptide S65Y66G67 fragment. Folding of the tertiary structure is fast, though chromophore formation and fluorescence is not observed until 90 min to 4 h after protein synthesis [16–18]. The folding of GFP exhibits hysteresis due to the decreased flexibility of the chromophore vs. its immature analogue [19–21], and the compaction of the β-barrel upon chromophore formation [22]. All signs of hysteresis disappear in mutants that do not form the chromophore [23]. Single-molecule fluorescence [24], mechanical [25] studies, and simulations thereof [26], NMR [27], denaturing

* Corresponding author. Address: Box 5624, Connecticut College, 270 Mohegan Ave., New London, CT 06320, USA.
E-mail address: mzim@conncoll.edu (M. Zimmer).

0301-0104/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemphys.2013.11.015

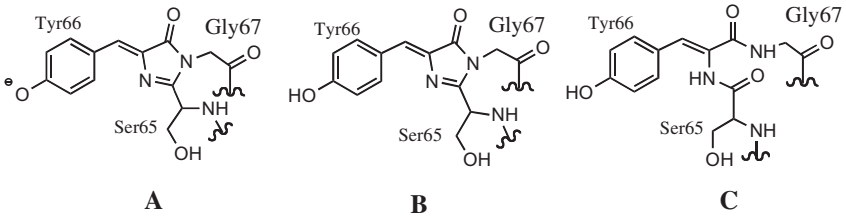

Fig. 1. Anionic (A) and neutral (B) forms of the chromophore. Neutral precyclized (immature) form (C).

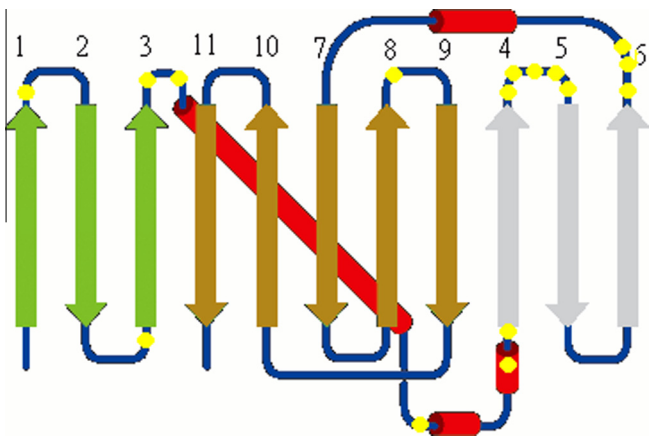

Fig. 2. Splay representation of GFP. The β sheet numbering is taken from the PDB, and is used throughout the paper. Conserved residues in the lids are indicated by yellow circles. The β sheets are colored by their unfolding behavior. The β-sheets unfold in groups. In the primary unfolding mechanism the brown C-terminus strands are the first to unfold, while in secondary pathways the green N-terminus strands have been observed to be the first to detach from the β barrel [56]. sheets β_{1-3}, β_{4-6} and β_{7-11} move as groups.

and renaturing experiments [28] and coarse-grained molecular simulations [29] have been used to examine folding in fluorescent proteins. Reddy et al. have combined the results from their coarse grained simulations with reported experimental observations [25,27,28,30] to propose a model for GFP folding [29]. It includes multiple pathways, passing through kinetic and equilibrium intermediates as well as misfolded structures.

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

2. Methods

2.1. Structure preparation of immature wild-type FPs and conformational searches

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

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

2.2. Structural conservation

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

2.3. Molecular dynamics

The coordinates of the *A. victoria* GFP crystal structure (1gfl) [63] were obtained from the Protein Data Bank (PDB) [31] and prepared as described above. 15000 MC steps were taken in each search. Structures within 50 kJ/mol of the lowest energy minimum were kept, and a usage directed method [60] was used to select structures for subsequent MC steps. Structures found in the conformational search were considered unique if the least squared superimposition of equivalent non-hydrogen atoms found one or more pairs separated by 0.25 Å or more. The lowest energy structure obtained in the search was further subjected to a 5000 step large scale low mode conformational search [64,65].

The final structures obtained from the fully minimized pdb structures and the conformational searches were used to initiate molecular dynamics (MD). MD simulations were carried out in the NPT ensemble at 300 K and 1 bar with 1.5 fs steps using Desmond [66]. All molecular dynamics calculations used the OPLS_2005 force field and SHAKE constrained hydrogens. 10418 structures were sampled in each 50 ns MD simulation. Each structure was in an orthorhombic simulation box of 0.15 M NaCl and SPC waters [67], with a 10 Å solvent buffer between the protein