Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

X-ray diffraction analysis and molecular-replacement solution of the cyan fluorescent protein dsFP483

Meitian Wang, Hetal N. Patel and Rebekka M. Wachter*

Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona, USA

Correspondence e-mail: rwachter@asu.edu

Received 25 August 2005
Accepted 10 September 2005
Online 30 September 2005

A novel cyan fluorescent protein, dsFP483 from the coral *Discosoma striata*, has been crystallized. Diffraction data were collected to 2.1 Å and processed in space group *C*2. Molecular-replacement methods were applied using the closely related red fluorescent protein DsRed as a search model. The asymmetric unit appears to contain six protein molecules (1.5 tetramers), five of which (83%) could be located by the molecular-replacement searches.

1. Introduction

The cyan fluorescent protein dsFP483 (*Discosoma striata* fluorescent protein) is a member of the green fluorescent protein-like (GFP-like) family of proteins (Matz *et al.*, 1999). DsFP483 emits bright visible light around 483 nm, in the cyan region of the spectrum, significantly blue-shifted from the 509 nm green emission of wild-type GFP (Tsien, 1998). Although GFP was originally isolated from the Pacific Northwest jellyfish, a large number of coloured and fluorescent proteins distantly related to GFP have recently been cloned from other marine organisms (Verkhusha & Lukyanov, 2004). Many of them, including dsFP483, are derived from non-bioluminescent reef-building corals of the class Anthozoa and lend these organisms their colourful appearance. However, all members of the GFP-like family of proteins appear to have an 11-stranded β-barrel fold with an internal helix that bears a protein-derived chromophore (Ormo *et al.*, 1996; Remington *et al.*, 2005; Wall *et al.*, 2000; Yang *et al.*, 1996; Yarbrough *et al.*, 2001). In these proteins, chromophore biogenesis proceeds spontaneously once the native protein fold is obtained and entails the autocatalytic modification of three internal amino-acid residues (Rosenow *et al.*, 2004; Tsien, 1998). In all GFP-like proteins characterized to date, a large part of the chromophore π-system is derived from a tyrosine residue. In GFP itself, the mature chromophore consists of a benzylidene imidazolinone entity, whereas in some of the homologues additional covalent modifications lead to a more extended chromophore π-system. Therefore, the fluorescence emission observed in this family of proteins may originate from one of several chemically related chromophores whose colours range from cyan, green and yellow to the red and far-red regions of the spectrum (Remington, 2002). In addition, the optical properties of these chromophores are modulated by the particular protein environment around the light-emitting entity in each homologue.

Biotechnological applications of fluorescent proteins are numerous and diverse and range from their use as fusion tags and reporter genes to a variety of biosensor applications (Tsien, 1998; Verkhusha & Lukyanov, 2004). Cyan fluorescent proteins engineered from GFP by molecular-biology techniques have been used extensively in fluorescence resonance energy-transfer (FRET) experiments (Rizzo *et al.*, 2004). The naturally occurring cyan-emitting protein dsFP483 is produced by the coral *D. striata*, an organism that exhibits blue–green stripes on its oral disc (Matz *et al.*, 1999). Though dsFP483 shares only about 25% sequence identity with GFP, it is more closely related to the *Discosoma* red fluorescent protein drFP583 (DsRed; 55% sequence identity). In GFP, the three chromophore-forming residues consist of Ser-Tyr-Gly, whereas in both the cyan and the red *Discosoma* proteins the equivalent residues consist of Gln-Tyr-Gly. While

DsRed undergoes additional covalent modifications that lead to red-colour emission from a green intermediate stage (Yarbrough *et al.*, 2001), the cyan protein is thought to contain a chromophore π-system similar to that of GFP. Yet dsFP483 is not green-fluorescent, but instead exhibits a strong absorption band at 437 nm with fluorescence emission at 483 nm. To better understand the structural basis that underlies the cyan colour of dsFP483, we have solved the X-ray structure of this protein by molecular replacement.

2. Expression, purification and crystallization of dsFP483

DsFP483, six-His-tagged at the N-terminus, was expressed in *Escherichia coli* JM109(DE3) using the expression plasmid pQE30 (Qiagen). Cultures were induced with 2 m*M* IPTG at an optical density OD_{600} of 0.6 and expression was allowed to proceed for 4 h at 298 K. The cyan protein was purified by affinity chromatography using Ni–NTA Superflow resin (Quiagen) and its concentration was determined by 280 nm absorbance. Crystals of dsFP483 were grown using the hanging-drop vapor-diffusion method by the addition of 2 µl protein (10 mg ml^{-1}) to 2 µl reservoir solution. Mother liquor consisted of 100 m*M* HEPES pH 7.9, 300 m*M* NaCl, 200 m*M* calcium acetate and 16%(*v*/*v*) PEG 4000. Diffraction-quality crystals grew at 277 K within 7–8 d. Crystals were transferred to a cryoprotectant solution [mother liquor plus 20%(*v*/*v*) glycerol] and stored at 77 K in a liquid-nitrogen dewar until data collection.

3. Data collection, processing and molecular-replacement solutions

Diffraction data were collected at 100 K using an ADSC Quantum 4R detector at Advanced Light Source beamline 5.0.3 (Lawrence Berkeley National Laboratory). Indexing and integration were carried out with *DENZO* (Otwinowski & Minor, 1997) and the data were scaled and merged with *SCALEPACK*. Data processing was carried out using a *C*-centered monoclinic cell (*C*2; $a = 111.2$, $b = 78.27$, $c = 188.6$ Å, $\beta = 91.35^\circ$). The overall R_{sym} was 0.070 (Table 1).

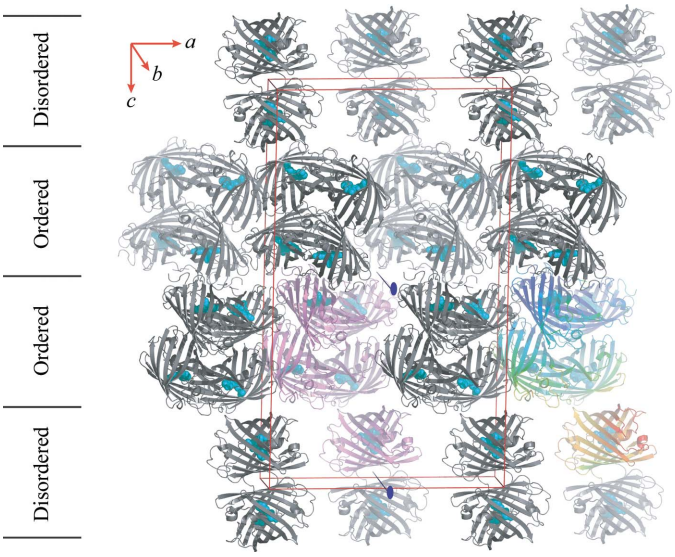

Figure 1
Molecular packing of dsFP483 in the *C*2 cell. Protein packing in the crystal may be thought of as a series of two-dimensional layers stacked upon each other along the crystallographic *c* axis. Protein molecules are represented as cartoon figures and the chromophores are represented as space-filling CPK models in cyan. The *C*2 cell is shown in red, together with two twofold crystallographic rotation axes. Molecules drawn in violet are the modelled contents of one asymmetric unit in the *C*2 cell.

Table 1
Crystallographic statistics for dsFP483.

Values in parentheses refer to the high-resolution shell.

Space group	*C*2 (No. 5)
Unit-cell parameters	
a (Å)	111.2
b (Å)	78.27
c (Å)	188.6
α (°)	90.00
β (°)	91.35
γ (°)	90.00
Data collection	
Detector	ADSC Quantum 4R
Wavelength (Å)	1.0000
Resolution (Å)	2.1
High-resolution shell (Å)	2.18–2.10
Total observations	636719
Unique reflections	90778
Redundancy	7.0 (5.1)
$\langle I/\sigma(I)\rangle$	21.8 (5.4)
Completeness (%)	94.8 (67.5)
R_{sym}† (%)	7.0 (35.9)

† $R_{sym} = \sum_h \sum_i I_i(h) - \langle I(h)\rangle|/\sum_h \sum_i I_i(h)$, where $I_i(h)$ is the *i*th intensity measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$.

Based on the high sequence identity with DsRed, the quarternary structure of dsFP483 was predicted to be tetrameric with pseudo-222 point-group symmetry. Hence, the DsRed tetramer was used as a search model (PDB code 1g7k; Yarbrough *et al.*, 2001) in molecular-replacement calculations carried out with the program *MOLREP* (Vagin & Teplyakov, 1997). Using this method, the position of one dsFP483 tetramer was easily located in the *C*2 cell. Molecular packing was visually inspected with the program *XFIT* (McRee, 1999) and the tetramer was found to form two-dimensional layers perpendicular to the crystallographic *c* axis, with enough empty space to accommodate another tetramer-based layer. The spacing of layers in the *C*2 cell indicated that each asymmetric unit should contain six protein molecules, *i.e.* one tetramer and one dimer, with the dimer part of a tetrameric structure generated by crystallographic twofold symmetry (Fig. 1; molecules occupying one asymmetric unit are coloured violet and the *C*2 cell is drawn in red). With six protein molecules in the asymmetric unit, the Matthews coefficient was calculated to be 2.6 Å^3 Da^{-1} and the solvent content 52.1%. Additional molecular-replacement searches were carried out to locate the presumably missing dimer, this time using only one protein chain of DsRed as a search model. This search was able to identify the location of a fifth dsFP483 molecule in the asymmetric unit, although with a relatively weak signal. The position of the sixth monomer could not be identified by molecular-replacement methods.

We considered the possibility of hemihedral crystal twinning, since the *C*2 cell has a β angle close to 90°. However, the observed cumulative intensity distribution was normal, with no indication of a sigmoidal profile that would be consistent with twinning. As a further test for twinning, the ratio of the average of the squared intensities to the square of the average intensity ($\langle I^2\rangle/\langle I\rangle^2$) was calculated using the acentric reflections (Yeates, 1997). This ratio was determined to be 2.22, close to the typical value of 2.0 expected for untwinned data, and rather dissimilar to the typical value of 1.5 expected for perfectly twinned data. Based on these tests, crystal twinning was judged to be highly unlikely. Therefore, a reasonable interpretation of the diffraction data is that the *C*2 asymmetric unit contains a total of six fluorescent protein monomers, one of them positionally disordered in the crystal (Fig. 1). A detailed structural analysis of the dsFP483 chromophore and its environment will need to await the construction

of a more complete crystallographic model and the utilization of a refinement strategy that takes disordered regions of the asymmetric unit into account (Winn *et al.*, 2003).

This work was supported by a grant from the National Science Foundation (MCB-0213091) to RMW. Crystallographic diffraction data were collected at beamline 5.0.3, Advanced Light Source, Lawrence Berkeley National Laboratory, which is supported by the US Department of Energy.

References

McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L. & Lukyanov, S. A. (1999). *Nature Biotechnol.* **17**, 969–973.
Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. & Remington, S. J. (1996). *Science*, **273**, 1392–1395.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Remington, S. J. (2002). *Nature Biotechnol.* **20**, 28–29.
Remington, S. J., Wachter, R. M., Yarbrough, D. K., Branchaud, B. P., Anderson, D. C., Kallio, K. & Lukyanov, K. A. (2005). *Biochemistry*, **44**, 202–212.
Rizzo, M. A., Springer, G. H., Granada, B. & Piston, D. W. (2004). *Nature Biotechnol.* **22**, 445–449.
Rosenow, M. A., Huffman, H. A., Phail, M. E. & Wachter, R. M. (2004). *Biochemistry*, **43**, 4464–4472.
Tsien, R. Y. (1998). *Annu. Rev. Biochem.* **67**, 509–544.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Verkhusha, V. & Lukyanov, K. A. (2004). *Nature Biotechnol.* **22**, 289–296.
Wall, M. A., Socolich, M. & Ranganathan, R. (2000). *Nature Struct. Biol.* **7**, 1133–1138.
Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
Yang, F., Moss, L. G. & Phillips, G. N. (1996). *Nature Biotechnol.* **14**, 1246–1251.
Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V. & Remington, S. J. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 462–467.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.