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

Structure of a fluorescent protein from *Aequorea victoria* bearing the obligate-monomer mutation A206K

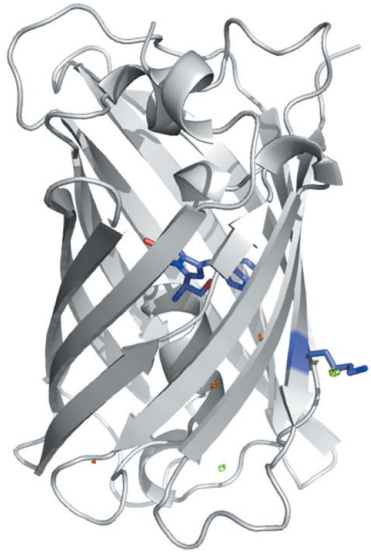
David von Stetten,[a] Marjolaine Noirclerc-Savoye,[b,c,d] Joachim Goedhart,[e] Theodorus W. J. Gadella Jr[e] and Antoine Royant[a,b,c,d]*

[a]Structural Biology Group, European Synchrotron Radiation Facility, 6 Rue Jules Horowitz, 38043 Grenoble, France, [b]CNRS, Institut de Biologie Structurale Jean-Pierre Ebel, 41 Rue Jules Horowitz, 38027 Grenoble, France, [c]CEA, Institut de Biologie Structurale Jean-Pierre Ebel, 41 Rue Jules Horowitz, 38027 Grenoble, France, [d]Université Joseph Fourier, Institut de Biologie Structurale Jean-Pierre Ebel, 41 Rue Jules Horowitz, 38027 Grenoble, France, and [e]Swammerdam Institute for Life Sciences, Section of Molecular Cytology, van Leeuwenhoek Centre for Advanced Microscopy, University of Amsterdam, Amsterdam, The Netherlands

Correspondence e-mail: antoine.royant@ibs.fr

Received 22 April 2012
Accepted 24 June 2012

PDB Reference: mTurquoise, 4ar7.



The green fluorescent protein (GFP) from the jellyfish *Aequoria victoria* has been shown to dimerize at high concentrations, which could lead to artefacts in imaging experiments. To ensure a truly monomeric state, an A206K mutation has been introduced into most of its widely used variants, with minimal effect on the spectroscopic properties. Here, the first structure of one of these variants, the cyan fluorescent protein mTurquoise, is presented and compared with that of its dimeric version mTurquoise-K206A. No significant structural change is detected in the chromophore cavity, reinforcing the notion that this mutation is spectroscopically silent and validating that the structural analysis performed on dimeric mutants also applies to monomeric versions. Finally, it is explained why cyan versions of GFP containing the Y66W and N146I mutations do not require the A206K mutation to prevent dimerization at high concentrations.

1. Introduction

Green fluorescent protein (GFP) is a protein from the glowing jellyfish *Aequorea victoria* which became world-famous after it was realised that it could serve as a genetically encoded source of fluorescence. Several observations, including spectroscopic changes upon concentration increase (Ward *et al.*, 1982), the translational diffusion coefficient observed by fluorescence correlation microscopy (Terry *et al.*, 1995) and the presence of a crystallographic dimer in GFP crystals (Yang *et al.*, 1996), suggested that the protein is monomeric at low concentrations but dimerizes at high concentrations. This notion was finally proved by analytical ultracentrifugation experiments (Phillips, 1997), resulting in a dissociation constant of $\sim$100 μM.

The distinct properties of monomeric and dimeric GFP could advantageously be used to image homo-oligomerization or clustering processes of GFP-fused proteins (De Angelis *et al.*, 1998), but it was clear that the dimerization property could lead to artefacts in imaging experiments by creating unwanted clustering events and that a truly monomeric GFP was highly desirable. Zacharias and coworkers examined the structure of dimeric GFP (Yang *et al.*, 1996) and identified a hydrophobic patch centred on Ala206, Leu221 and Phe223 of each monomer. By mutating one or several of these residues to positively charged amino-acid side chains, they succeeded in efficiently removing the interaction between monomers (Zacharias *et al.*, 2002). When the A206K mutation was introduced into yellow fluorescent protein (YFP), the dissociation constant increased from 0.11 to 74 mM. From then on, the letter 'm' was added to the acronyms of obligate-monomer versions of fluorescent proteins. The A206K mutation gained popularity among *A. victoria* fluorescent proteins and is present in the cyan variants m(E)CFP (Zacharias *et al.*, 2002), mTurquoise (Goedhart *et al.*, 2010), mTurquoise2 (Goedhart *et al.*, 2012) and mCerulean3 (Markwardt *et al.*, 2011), the green variant m(E)GFP (Kremers *et al.*, 2011; Zacharias *et al.*, 2002) and the yellow variants mCitrine (Griesbeck *et al.*, 2001; Shaner *et al.*, 2005) and mVenus (Nagai *et al.*, 2002; Kremers *et al.*, 2006). The mutation has been shown to be spectroscopically silent (Zacharias *et al.*, 2002; Kremers *et al.*, 2006), but the question of whether this would have an effect on the environment of the chromophore remained open in the absence of a structure of a fluorescent protein containing the A206K mutation.

Cyan fluorescent proteins (CFPs) contain the critical Y66W mutation in the chromophore (replacement of the central tyrosine by a tryptophan) that is responsible for the characteristic cyan colour of fluorescence emission. Here, we report the high-resolution X-ray structure of mTurquoise, a CFP with the A206K mutation, which represents the first structure of a truly monomeric fluorescent protein from the jellyfish *A. victoria*.

2. Materials and methods

2.1. Protein expression and purification

The mTurquoise coding sequence, originally present in a pRSET vector (Goedhart *et al.*, 2010), was transferred into pQE60-Cerulean (Lelimousin *et al.*, 2009) using the *Nco*I and *Bsr*GI restriction sites to replace Cerulean. The recombinant protein was expressed in *Escherichia coli* BL21 CodonPlus (DE3) RIL cells (Stratagene) in autoinduction medium at 290 K for 24 h. The cells were lysed by sonication in the presence of 20 m*M* Tris pH 8.0 and 500 m*M* NaCl with EDTA-free protease inhibitors (Complete, Roche). The His-tagged protein was purified on an Ni–NTA Superflow column (Qiagen) and eluted with 100 m*M* imidazole in the aforementioned buffer. Fractions containing purified proteins were pooled, dialyzed against 20 m*M* Tris pH 8.0 and concentrated to 38 mg ml^{-1}.

2.2. Crystallization, *in crystallo* spectroscopy, data collection and structure determination

mTurquoise was crystallized using the hanging-drop method at 293 K at a concentration of 6 mg ml^{-1} in 15% PEG 8000, 100 m*M* $MgCl_2$, 100 m*M* HEPES pH 6.5. Crystals grew more slowly and were smaller than those of mTurquoise-K206A (Goedhart *et al.*, 2012); they typically grew in 30 d (compared with 10 d for mTurquoise-K206A) as long thin needles of maximum width 20 µm. UV–visible absorption and fluorescence emission spectra of mTurquoise crystals were recorded at 100 K at the ESRF Cryobench laboratory (Royant *et al.*, 2007) and proved to be indistinguishable from those of mTurquoise-K206A crystals (Goedhart *et al.*, 2012; Fig. 1). Helical data collection (Flot *et al.*, 2010) was performed at 100 K on beamline ID29 (wavelength 0.984 Å) of the European Synchrotron Radiation Facility, Grenoble (de Sanctis *et al.*, 2012). A 1.23 Å resolution data set was integrated with *XDS* (Kabsch, 2010) and scaled and merged with *SCALA* (Winn *et al.*, 2011). The 1.47 Å resolution structure of mTurquoise-K206A (PDB entry 2ye0; Goedhart *et al.*, 2012) was used as a starting model for refinement after all anisotropic *B* factors had been reset to isotropic *B* factors and was refined with *REFMAC*5 (Murshudov *et al.*, 2011) with anisotropic temperature factors. *Coot* (Emsley *et al.*, 2010) was used for model inspection and building of the side chain of Lys206. Data-reduction and structure-refinement statistics can be found in Table 1. The structure and structure-factor amplitudes for mTurquoise have been deposited in the Protein Data Bank (http://www.pdb.org/) as PDB entry 4ar7. The Fourier difference map was calculated using the programs *CAD*, *SCALEIT* and *FFT* from the *CCP*4 suite (Winn *et al.*, 2011). Structure figures were prepared with *PyMOL* (http://www.pymol.org/).

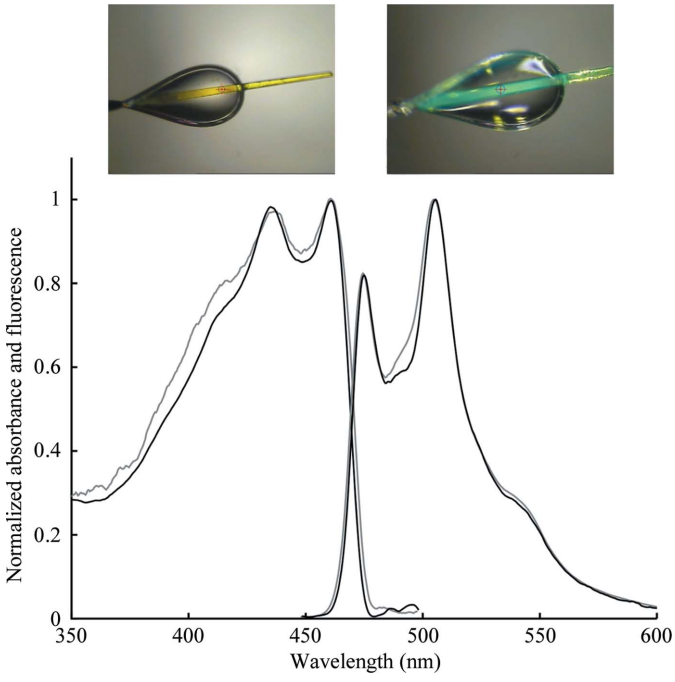


Figure 1
Spectroscopic properties of mTurquoise crystals. UV–visible absorption and fluorescence emission spectra recorded at 100 K for mTurquoise (black) and mTurquoise-K206A (Goedhart *et al.*, 2012; grey) crystals. Insets show photographs of the crystal used for structure determination, illuminated by only the back light of the beamline (left inset; transmitted light) and by both back and front lights (right inset; transmitted and reflected light). The right inset shows that the front light can generate detectable fluorescence of the outer layers of the crystal.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data processing	
Beamline	ID29, ESRF
Wavelength (Å)	0.984
Temperature (K)	100
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 51.45, b = 62.56, c = 69.54
Resolution range (Å)	51.5–1.23 (1.30–1.23)
Observations	278479
Unique reflections	62924
R_{merge} (%)	4.8 (45.1)
$\langle I/\sigma(I)\rangle$	13.5 (2.8)
Multiplicity	4.4 (4.3)
Completeness (%)	95.9 (90.3)
Refinement	
R_{cryst}/R_{free} (%)	14.3/16.6
No. of atoms	2039
Protein	1825
Chromophore	23
Water	191
Average B factor (Å^2)	
Overall	18.7
Protein	17.7
Chromophore	10.6
Water	28.8
R.m.s. deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.9
Ramanchandran plot (No. of residues)	
Preferred regions	209 [96.3%]
Allowed regions	7 [3.2%]
Outliers	1 [0.5%]
PDB code	4ar7

3. Results and discussion

3.1. Comparison of the mTurquoise and mTurquoise-K206A structures

To verify the presence of the mutated residue in the crystal structure, we first calculated a Fourier difference map between the published 1.47 Å resolution data set of mTurquoise-K206A (PBD

entry 2ey0; Goedhart *et al.*, 2012) and the 1.23 Å resolution data set of mTurquoise using the calculated phases of the mTurquoise-K206A structure. The strongest peak in the resulting map is a positive peak located next to the C^{β} atom of Ala206, showing the presence of C^{γ} of the lysine (Fig. 2*a*). The next strongest peaks are located at the radiation-sensitive S atoms of Cys48 and Met218, indicating different absorbed X-ray doses for the data collections. The difference map suggests that structural changes between mTurquoise-K206A and mTurquoise are very limited. The structure of mTurquoise was refined to $R_{\rm cryst}$ = 14.3% and $R_{\rm free}$ = 16.6%. It superimposed very well on the structure of mTurquoise-K206A, with a root-mean-square deviation of only 0.107 Å on the superpositioning of all C^{α} atoms. A close look at the active site shows that all of the residues surrounding the chromophore are almost perfectly superimposed (Fig. 2*b*). The largest distance change (0.16 Å) is found between the chromophore and Ile146, which has been shown to be significantly mobile in all CFPs (Goedhart *et al.*, 2012; Lelimousin *et al.*, 2009). All other distance changes are less than 0.06 Å, which is well below the estimated overall coordinate error based on maximum likelihood (0.24 Å) and strongly suggests that the environment of the chromophore has not changed upon the K206A mutation. Thus, this validates that structural analysis of CFPs with an alanine at position 206 (Bae *et al.*, 2003; Lelimousin *et al.*, 2009; Goedhart *et al.*, 2012) is relevant to the equivalent CFPs with a lysine at position 206.

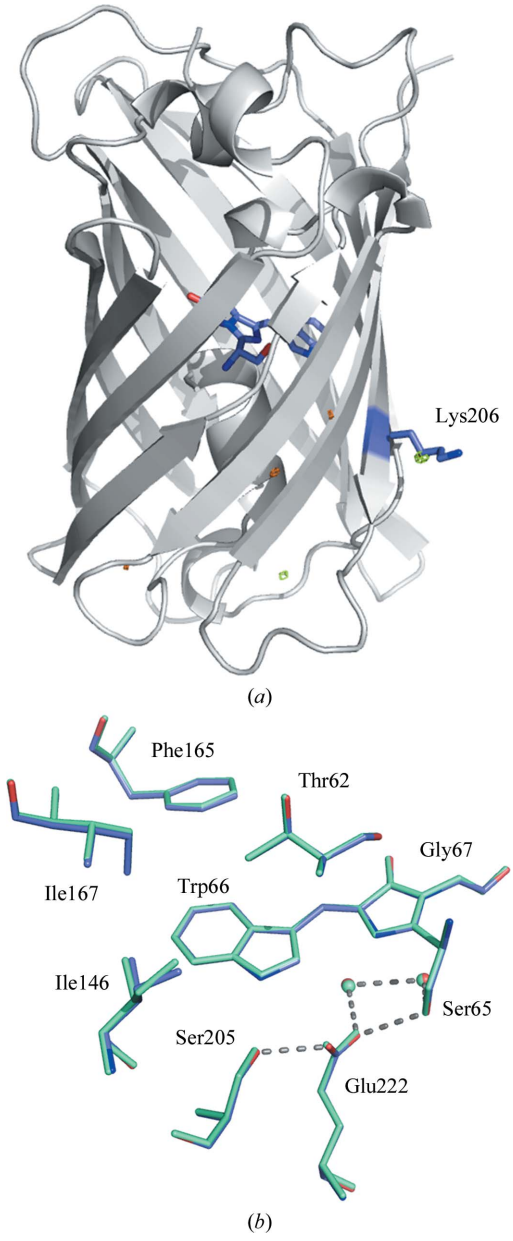


Figure 2
Structure of mTurquoise. (*a*) Cartoon representation of the mTurquoise structure with the tryptophan-based chromophore shown in blue. A Fourier difference map between a data set from mTurquoise-K206A (PDB entry 2ye0) and mTurquoise is superimposed on the structure at a contour level of ±6σ (green, positive; orange, negative). The strongest peak in the map is located at the C^{γ} atom of Lys206, suggesting that there is no significant change between the two structures. (*b*) Superimposition of the chromophore and surrounding residues in mTurquoise-K206A (cyan) and mTurquoise (blue). Changes in distances to the chromophore are well within experimental error, with the largest change being at the side chain of Ile146, a residue observed to be significantly mobile in all CFP structures.

3.2. Implications of the N146I mutation for CFP dimer formation

The structure of the physiological dimer of GFP (Fig. 3*a*) can be inferred from the first structure of wild-type GFP (PDB entry 1gfl) obtained from crystals in space group $P4_12_12$ with two molecules in the asymmetric unit (Yang *et al.*, 1996). It has recently been demonstrated using fluorescence anisotropy measurements that in contrast to other *A. victoria* mutants the cyan fluorescent protein ECFP does not dimerize at concentrations up to millimolar (Espagne *et al.*, 2011). This could be explained by the presence of one of the two characteristic mutations of CFP, N146I, which was initially introduced to restore a significant level of fluorescence to the Y66W mutant of wild-type GFP (*i.e.* replacement of the tyrosine of the chromophore by a tryptophan). Asn146 is positioned on the seventh strand of the β-barrel of GFP with its side chain exposed to the solvent, while the neighbouring residue Tyr145 has its side chain oriented towards the chromophore inside the protein (Fig. 3*b*; Yang *et al.*, 1996). A detailed analysis of two CFP structures, those of ECFP and Cerulean, showed that the N146I mutation results in the rotation of the seventh strand of the β-barrel, exposing the side chain of Tyr145 to the solvent and sheltering the hydrophobic side chain of Ile146 inside the protein in the vicinity of the chromophore (Bae *et al.*, 2003; Lelimousin *et al.*, 2009; Fig. 3*b*). As a consequence, if we build a putative CFP dimer by aligning a monomer on each of the two GFP molecules (the blue trace in Fig. 3*b*) it becomes apparent that the Tyr145 residues in each monomer clash with each other, hindering the formation of a CFP dimer in a GFP-like arrangement.

3.3. Implications of the A206K mutation for dimer formation in other *A. victoria* fluorescent proteins

The availability of the mTurquoise structure makes it possible to test the structural implications of the A206K mutation for dimer formation. Fig. 3(*c*) shows the proximity of Lys206 in the molecules of a putative mTurquoise dimer. Each of these lysines brings a positive charge into the close vicinity (between 1.9 and 2.6 Å) of the two residues of the hydrophobic patch forming the core of the dimer interface. Obviously, their presence makes formation of the dimer impossible at any concentration. Fig. 3(*d*) summarizes all interactions that occur at the dimer interface in mTurquoise and GFP. Obviously, one should not rule out the formation of dimers at very high concentrations through another interface elsewhere on the protein surface. Finally, the dimer disruption by Lys206 observed in the structure of a CFP is likely to be applicable to the structures of other CFPs, such as ECFP, Cerulean, SCFP3A and mTurquoise2 (Goedhart *et al.*, 2012; Lelimousin *et al.*, 2009), and also to those of mEGFP, mVenus and mCitrine based on the structures of their dimeric forms

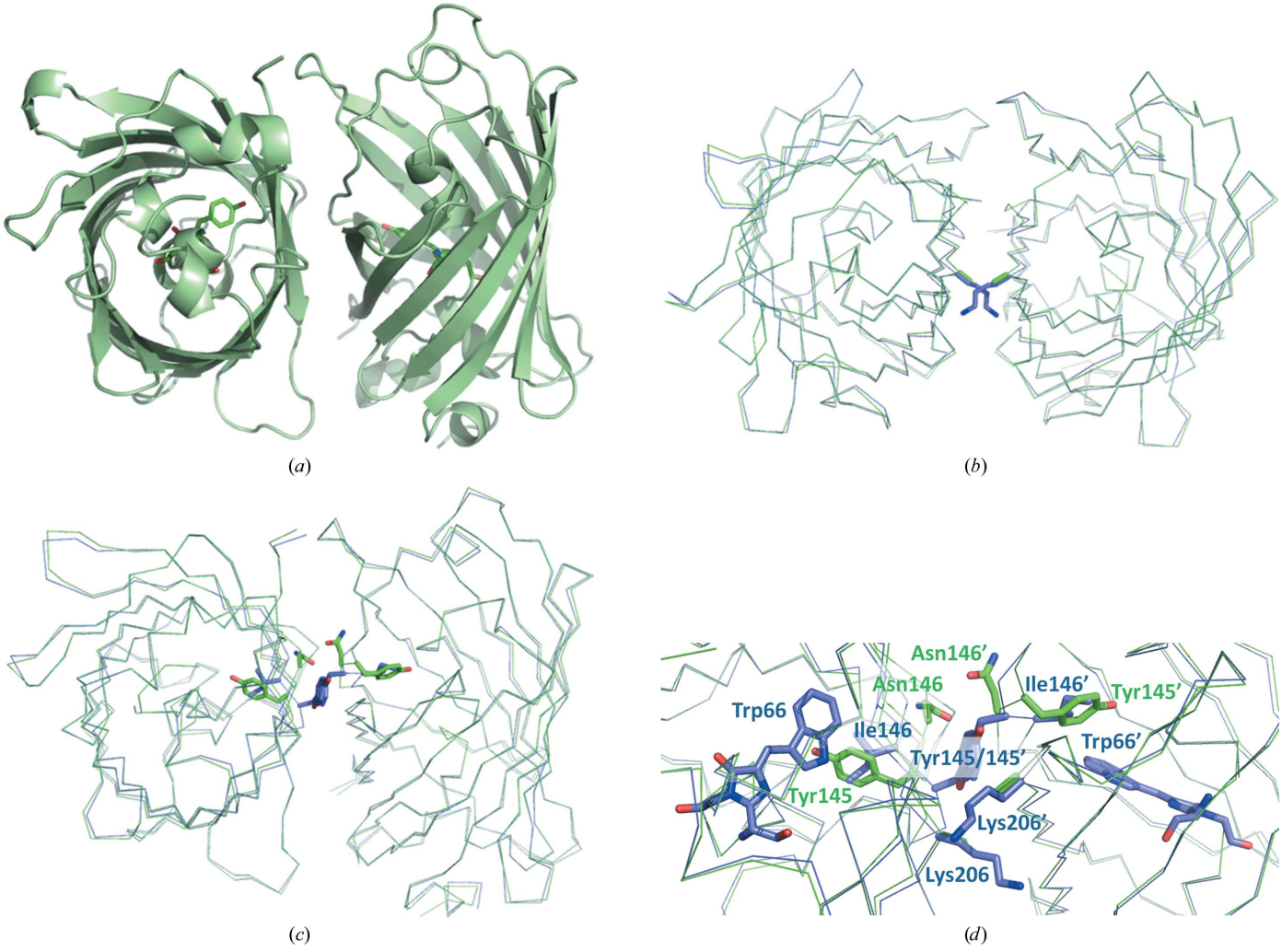


Figure 3
Structural comparison of a wild-type GFP dimer and a putative mTurquoise dimer. (*a*) Cartoon representation of the crystallographic dimer observed in the first published structure of wild-type GFP (Yang *et al.*, 1996). (*b*) Ribbon representation of the GFP dimer and of a virtual mTurquoise dimer constructed by aligning the structure of mTurquoise on each molecule of the GFP dimer. The overlap between the two side chains of Tyr145 in mTurquoise (and all other CFP mutants for which structures have been determined to date) shows why CFP mutants which carry the N146I mutation do not show oligomerization even at high concentration. (*c*) The same view as in (*b*) with residues 206 represented. The two positive charges borne by Lys206 would induce electrostatic repulsion in any dimer analogous to the GFP dimer even if the N146I mutation were not present. (*d*) Close-up of the dimer interface. GFP residues are labelled in green and mTurquoise residues are labelled in blue.

(Griesbeck *et al.*, 2001; Nagai *et al.*, 2002; Royant & Noirclerc-Savoye, 2011).

4. Conclusion

By solving the structure of mTurquoise from *A. victoria*, which has a lysine residue at position 206, we have shown that the A206K mutation does not affect the structure of the chromophore environment. This result is likely to be applicable to other mutants of *A. victoria* GFP. Additionally, we provide a structural explanation for the different oligomerization behaviour of CFPs compared with other fluorescent proteins from *A. victoria*.

We thank the ESRF Structural Biology beamline staff for assistance and advice during data collection. The Partnership for Structural Biology is acknowledged for access to the RoBioMol and Cryobench platforms.

References

Bae, J. H., Rubini, M., Jung, G., Wiegand, G., Seifert, M. H. J., Azim, M. K., Kim, J.-S., Zumbusch, A., Holak, T. A., Moroder, L., Huber, R. & Budisa, N. (2003). *J. Mol. Biol.* **328**, 1071–1081.
De Angelis, D. A., Miesenböck, G., Zemelman, B. V. & Rothman, J. E. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 12312–12316.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Espagne, A., Erard, M., Madiona, K., Derrien, V., Jonasson, G., Lévy, B., Pasquier, H., Melki, R. & Mérola, F. (2011). *Biochemistry*, **50**, 437–439.
Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S. & Mitchell, E. (2010). *J. Synchrotron Rad.* **17**, 107–118.
Goedhart, J., van Weeren, L., Hink, M. A., Vischer, N. O., Jalink, K. & Gadella, T. W. (2010). *Nature Methods*, **7**, 137–139.
Goedhart, J., von Stetten, D., Noirclerc-Savoye, M., Lelimousin, M., Joosen, L., Hink, M. A., van Weeren, L., Gadella, T. W. & Royant, A. (2012). *Nature Commun.* **3**, 751.
Griesbeck, O., Baird, G. S., Campbell, R. E., Zacharias, D. A. & Tsien, R. Y. (2001). *J. Biol. Chem.* **276**, 29188–29194.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kremers, G. J., Gilbert, S. G., Cranfill, P. J., Davidson, M. W. & Piston, D. W. (2011). *J. Cell Sci.* **124**, 157–160.
Kremers, G. J., Goedhart, J., van Munster, E. B. & Gadella, T. W. (2006). *Biochemistry*, **45**, 6570–6580.
Lelimousin, M., Noirclerc-Savoye, M., Lazareno-Saez, C., Paetzold, B., Le Vot, S., Chazal, R., Macheboeuf, P., Field, M. J., Bourgeois, D. & Royant, A. (2009). *Biochemistry*, **48**, 10038–10046.
Markwardt, M. L., Kremers, G. J., Kraft, C. A., Ray, K., Cranfill, P. J., Wilson, K. A., Day, R. N., Wachter, R. M., Davidson, M. W. & Rizzo, M. A. (2011). *PLoS One*, **6**, e17896.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K. & Miyawaki, A. (2002). *Nature Biotechnol.* **20**, 87–90.
Phillips, G. N. Jr (1997). *Curr. Opin. Struct. Biol.* **7**, 821–827.
Royant, A., Carpentier, P., Ohana, J., McGeehan, J., Paetzold, B., Noirclerc-Savoye, M., Vernède, X., Adam, V. & Bourgeois, D. (2007). *J. Appl. Cryst.* **40**, 1105–1112.
Royant, A. & Noirclerc-Savoye, M. (2011). *J. Struct. Biol.* **174**, 385–390.
Sanctis, D. de *et al.* (2012). *J. Synchrotron Rad.* **19**, 455–461.
Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. (2005). *Nature Methods*, **2**, 905–909.
Terry, B. R., Matthews, E. K. & Haseloff, J. (1995). *Biochem. Biophys. Res. Commun.* **217**, 21–27.
Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W. & Reeves, S. C. (1982). *Photochem. Photobiol.* **35**, 803–808.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yang, F., Moss, L. G. & Phillips, G. N. Jr (1996). *Nature Biotechnol.* **14**, 1246–1251.
Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. (2002). *Science*, **296**, 913–916.