Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization of small proteins assisted by green fluorescent protein

Nobuhiro Suzuki,[a]‡ Masahiko Hiraki,[a] Yusuke Yamada,[a] Naohiro Matsugaki,[a] Noriyuki Igarashi,[a] Ryuichi Kato,[a] Ivan Dikic,[b] David Drew,[c] So Iwata,[c] Soichi Wakatsuki[a] and Masato Kawasaki[a]*

[a]Structural Biology Research Center, Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan, [b]Institute of Biochemistry II and Cluster of Excellence Frankfurt, Goethe University School of Medicine, Theodor-Stern-Kai 7, D-60590 Frankfurt am Main, Germany, and [c]Division of Molecular Biosciences, Membrane Protein Crystallography Group, Imperial College, London SW7 2AZ, England

‡ Present address: Protein Research Unit, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan.

Correspondence e-mail: masato.kawasaki@kek.jp

Received 18 May 2010
Accepted 16 August 2010

PDB References: GFP–ubiquitin-binding motif fusion, 3ai4; GFP–ubiquitin fusion, 3ai5.

The generation of crystal lattice contacts by proteinaceous tags fused to target proteins is an attractive approach to aid in the crystallization of otherwise intractable proteins. Here, the use of green fluorescent protein (GFP) fusions for this purpose is demonstrated, using ubiquitin and the ubiquitin-binding motif (UBM) of Y-family polymerase ι as examples. The structure of the GFP–ubiquitin fusion protein revealed that the crystal lattice was formed by GFP moieties. Ubiquitin was accommodated in the lattice through interactions with the peripheral loops of GFP. However, in the GFP–UBM fusion crystal UBM formed extensive interactions with GFP and these interactions, together with UBM dimerization, mediated the crystal packing. Interestingly, the tyrosine residues that are involved in mediating crystal contacts in both GFP–ubiquitin and GFP–UBM crystals are arranged in a belt on the surface of the β-barrel structure of GFP. Therefore, it is likely that GFP can assist in the crystallization of small proteins and of protein domains in general.

1. Introduction

Crystallization represents a major bottleneck in the determination of macromolecular structure by X-ray crystallography. In order to tackle challenging target proteins, several crystallization chaperones have been developed to aid in crystal lattice formation (for reviews, see Koide, 2009; Derewenda, 2010). For example, monoclonal antibodies have been used to obtain membrane-protein cocrystal structures by increasing the surface area available for crystal contacts (Koide, 2009; Derewenda, 2010). Designed ankyrin-repeat protein (DARPin) is a non-antibody scaffold protein with randomized surface residues that has also been successfully applied to a number of different proteins (Sennhauser & Grütter, 2008). However, the drawback of this approach is that these protein-crystallization chaperones must be specifically optimized, *e.g.* by natural immune systems or by *in vitro* library-screening techniques such as ribosome display (Koide, 2009).

An alternative approach to crystallization chaperones that does not require specific protein–protein interactions is protein fusion. The use of carrier proteins was introduced by Donahue and coworkers with the use of fibrinogen fragments fused to lysozyme (Donahue *et al.*, 1994). Recently, human β2-adrenergic and adenosine A2a G-protein-coupled receptors have been crystallized as fusions with T4 lysozyme (Cherezov *et al.*, 2007; Rosenbaum *et al.*, 2007; Jaakola *et al.*, 2008). This fusion technique does not require an exhaustive and costly screening of random mutations on the surface of proteins.

However, the presence of a flexible linker between the tag and target protein may allow conformational heterogeneity and thereby inhibit the formation of essential crystal contacts. The introduction of a short linker with a large affinity tag has proven to promote crystallization in a number of instances (Smyth *et al.*, 2003). The most prolific fusion protein is maltose-binding protein (MBP), with at least 23 MBP-fusion crystal structures of proteins or protein complexes deposited in the Protein Data Bank (PDB) to date (Moon *et al.*, 2010). Combination of this approach with surface-entropy reduction can significantly increase the success rate of crystallization (Moon *et al.*, 2010). Apart from MBP, the number of reported fusion-protein structures with other tags, such as glutathione *S*-transferase (GST), is still limited (Moon *et al.*, 2010).

We therefore questioned whether green fluorescent protein (GFP) could be used as a scaffold tag for protein crystallization. GFP is one of the most widely used fusion tags for tracking and visualizing proteins *in vivo*, chiefly because the fusion rarely affects the function of the protein partner and also because it requires no extra cofactors for the acquisition of fluorescence. In addition, there has been a surge in the use of GFP tagging to aid in the expression screening and purification of difficult-to-handle proteins. GFP has also been used to monitor the expression of both soluble and membrane proteins (Waldo *et al.*, 1999; Newstead *et al.*, 2007) in order to identify more highly expressed proteins or domains. Furthermore, fluorescence-detection size-exclusion chromatography (FSEC) using a GFP-fusion tag has been shown to help in the identification of crystallizable membrane proteins (Kawate & Gouaux, 2006).

Here, we report the successful application of the GFP-fusion method to aid in the crystallization of small proteins/domains with a short linker, using ubiquitin as an example of a small protein and a ubiquitin-binding motif (UBM) to represent a small protein domain. We have observed that tyrosine residues on the surface of the GFP β-barrel are arranged in a structural feature that we refer to as the 'tyrosine belt'. This tyrosine belt mediates protein–protein interactions in both GFP–ubiquitin and GFP–UBM fusion crystals. GFP therefore possesses the potential to become a more versatile tag to assist in the crystallization of small proteins or domains.

2. Materials and methods

2.1. Expression of GFP-fusion proteins

The coding sequence for residues 1–230 of yEGFP (GenBank accession No. U73901; Cormack *et al.*, 1997), a GFP with S65G and S72A mutations to enhance fluorescence (excitation maximum at 490 nm and emission maximum at 517 nm) and codon-optimized for yeast expression, was cloned into the *Nde*I–*Bam*HI sites of pET28a (Novagen) for expression in *Escherichia coli*. Subcloning of target-protein genes using the *Bam*HI restriction enzyme leaves a two-residue linker (Gly-Ser) between yEGFP(1–230) and the target protein. The fusion protein was expressed in *E. coli* BL21 (DE3) cells in LB medium by induction with 0.2 m*M* IPTG for 5 h at 310 K.

2.2. Purification, crystallization and structural analysis of the GFP–ubiquitin fusion protein

We attempted to express and crystallize a GFP–ubiquitin–UBM2 fusion protein in which yEGFP (1–230), ubiquitin (1–76) (GenBank accession No. X51703) and mouse polymerase ι UBM2 (668–717) (GenBank accession No. AF151691) were connected in tandem by Gly-Ser linkers. However, mass spectrometry revealed that the fusion protein was cleaved after the 74th residue of ubiquitin during expression and purification. Therefore, the cleaved GFP–ubiquitin (1–74) fusion construct, from which Gly75 and Gly76 of ubiquitin were missing, was the final crystallization target. The fusion protein was purified from the soluble fraction of the bacterial cell lysates by immobilized metal-affinity column chromatography (Ni–NTA Superflow, Qiagen) in a buffer consisting of 50 m*M* sodium phosphate, 300 m*M* NaCl and 20 m*M* imidazole. The N-terminal His_6 tag (MGSSHHHHHHSSGLVPR) was cleaved by thrombin on the column (295 K, 24 h), leaving two amino acids (GS) at the amino-terminus of GFP, and the eluted GFP–ubiquitin fusion protein was further purified by size-exclusion chromatography using Superdex 75 pg (GE Healthcare) in a buffer consisting of 20 m*M* Tris–HCl pH 8.0 and 100 m*M* NaCl. The fusion protein was concentrated to 20 mg ml^{-1} and subjected to our crystallization system (Hiraki *et al.*, 2006). The crystallization conditions for the GFP–ubiquitin fusion were screened using the Crystal Screen, Crystal Screen 2, PEG/Ion, PEG/Ion 2, Index (Hampton Research, USA) and Wizard I and II (Emerald BioSystems, USA) screening kits. The well volume was 180 µl and the drop consisted of 0.5 µl protein solution and 0.5 µl well solution. Crystals were cryoprotected by soaking them in mother liquor supplemented with 10%(*v*/*v*) ethylene glycol. The diffraction data sets were collected on beamline 13B1 at the National Synchrotron Radiation Research Center, Taiwan and NW12A at PF-AR, Tsukuba, Japan under cryogenic conditions. The structure of the fusion protein was determined by the molecular-replacement method using the program *MOLREP* (Vagin & Teplyakov, 1997). The coordinates of GFP (PDB code 1gfl; Yang *et al.*, 1996) and ubiquitin (PDB code 1ubq; Vijay-Kumar *et al.*, 1987) were used as search models. Cycles of refinement using *REFMAC*5 (Murshudov *et al.*, 1997) were alternated with manual model building using *Coot* (Emsley *et al.*, 2010).

2.3. Purification, crystallization and structural analysis of the GFP–UBM fusion protein

Mouse polymerase ι UBM2 (668–717) (GenBank accession No. AF151691) was fused with yEGFP (1–230). The fusion protein was purified using the same procedure as used for the GFP–ubiquitin fusion. The GFP–UBM2 fusion protein was concentrated to 40 mg ml^{-1} and screened using the crystallization conditions described above. Crystals were cryoprotected by soaking them in mother liquor supplemented with

10%(*v*/*v*) ethylene glycol. For structure determination, we performed molecular replacement using *MOLREP* (Vagin & Teplyakov, 1997) with GFP (PDB code 1gfl; Yang *et al.*, 1996) as the search model. This was followed by automatic model building using *ARP*/*wARP* (Perrakis *et al.*, 1999). The model was refined as described above. A structural homology search was performed using the *DALI* server (Holm *et al.*, 2008). As *DALI* requires the query proteins to be 30 residues or more in length and the UBM2 moiety in our structure was only 29 residues long, we added two Ala residues to the N-terminus of the electron-density map, enabling the *DALI* server to initiate calculation. The molecular-interface area in the crystal was calculated using the *PISA* server (Krissinel & Henrick, 2007). Structural figures were generated using *PyMOL* (http://www.pymol.org).

3. Results

3.1. GFP–ubiquitin fusion

As a first test case, we attempted to cocrystallize ubiquitin (8.5 kDa) with GFP. As it is known that the linker length is critical for successful crystallization of fusion proteins (Smyth *et al.*, 2003) and that the eight C-terminal residues (231–238) are invisible in most GFP-only crystal structures, we truncated yEGFP (Cormack *et al.*, 1997) after residue 230 to make a linker length of only two amino acids (Gly-Ser). The GFP–ubiquitin fusion protein readily crystallized in conditions containing PEG 3350 as a precipitant at a pH in the range 5.0–8.5, although the crystals took several weeks to grow. Crystals of GFP–ubiquitin are yellow under white light, green fluorescent under blue light (Fig. 1) and colourless under acidic conditions near pH 5.

The best diffraction data set, which was collected at 1.4 Å resolution, was obtained from a crystal grown in PEG/Ion 2 condition No. 48 [1%(*w*/*v*) Tryptone, 20%(*w*/*v*) PEG 3350]. The structure of the GFP–ubiquitin fusion was solved by the molecular-replacement method using GFP and ubiquitin as search models (Table 1). The Gly-Ser linker between GFP and ubiquitin was resolved in the crystal structure. There is no 'intramolecular' interaction between GFP and ubiquitin (Fig. 1*a*). Although the fusion appears to migrate as a monomer in size-exclusion chromatography, GFP forms a dimer in the crystal structure (with 930 Å^2 of solvent-accessible surface buried in the interface) *via* the β10 and β11 strands and the loop preceding the β7 strand. Crystal contacts are also facilitated by interaction between the two GFP dimers. More specifically, the β4, β5 and β9 strands of GFP interact with the N-terminal helix, the loop preceding β4 and the loop between β5 and β6 of the neighbouring GFP (Figs. 2*a* and 2*b*; interface area 373 Å^2). Notably, the Phe99 and Tyr182 residues of GFP are found in the centre of this GFP–GFP interface (Fig. 2*b*).

Only 20% (870 of 4269 Å^2) of the ubiquitin surface makes crystal contacts. The main interactions are observed in the β1 and β6 strands of ubiquitin, which interact with the β7–β8 and β9–β10 loops of GFP in the neighbouring molecule (with a surface area of 398 Å^2). Ubiquitin alone has been crystallized in several different crystal-packing systems (Vijay-Kumar *et al.*, 1987; Falini *et al.*, 2008). In these crystals, more than 25% of the surface of ubiquitin is engaged in crystal contacts: 1244 of 4789 Å^2 surface area (26%) is used for crystal packing in PDB entry 1ubq, 1327 of 4363 Å^2 (30%) in PDB entry

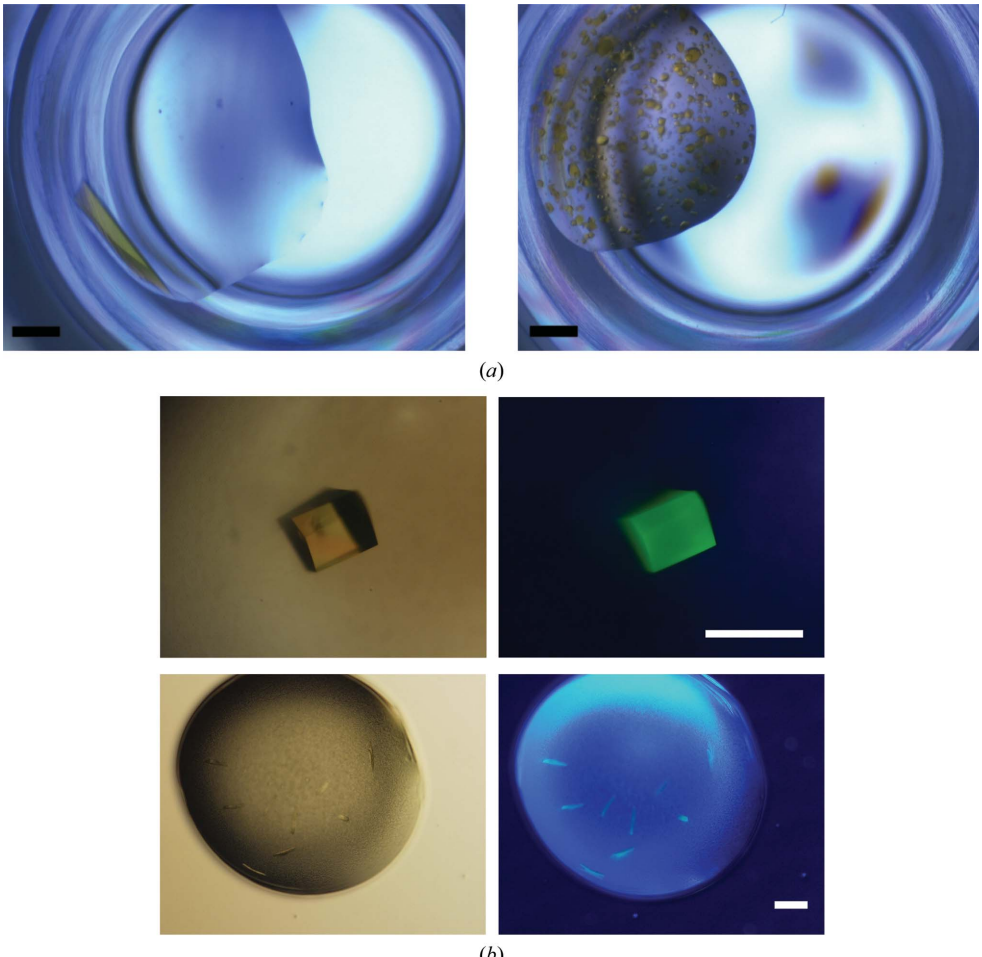

Figure 1
Crystals of GFP-fusion proteins. (*a*) Crystals of GFP–ubiquitin (left) and GFP–UBM2 (right). (*b*) Crystals of GFP–ubiquitin observed under white light (left panels) or blue LED light (right panels). In all cases the scale bar represents 0.3 mm.

3eec (chain *A*) and 2015 of 4231 Å^2 (48%) in PDB entry 3ehv (chain *A*). As one might expect, the GFP fusion reduces the surface area of the target required for crystal lattice formation. Ubiquitin has a canonical interaction surface centred around Ile44 that been shown to be involved in interaction with more than 16 ubiquitin-binding domains (Hurley *et al.*, 2006). In the GFP–ubiquitin fusion crystal, the Ile44 surface is not used for crystal packing but is exposed to the solvent (Fig. 2*c*).

3.2. GFP–UBM fusion

The ubiquitin-binding motif (UBM) is a recently discovered ubiquitin-binding domain that is present in Y-family polymerase ι and REV1, both of which contain two UBMs at the C-terminus (Bienko *et al.*, 2005). We attempted to crystallize the second UBM of mouse polymerase ι (UBM2; 6 kDa), with no success. Subsequently, we fused UBM2 (residues 668–717) to the C-terminus of yEGFP (1–230) using a Gly-Ser linker. The GFP–UBM2 fusion protein crystallized in several conditions containing either ammonium sulfate or sodium acetate as precipitant. Four conditions resulted in crystals with a yellowish colour: Crystal Screen No. 4 (2.0 *M* ammonium sulfate, 0.1 *M* Tris–HCl pH 8.5), Crystal Screen 2 No. 34 (1.0 *M* sodium acetate, 0.1 *M* HEPES pH 7.5, 0.05 *M* cadmium sulfate), Index No. 5 (0.1 *M* HEPES pH 7.5, 2.0 *M* ammonium sulfate) and Index No. 6 (0.1 *M* Tris–HCl pH 8.5, 2.0 *M* ammonium sulfate). The optimal data set was collected to 1.6 Å resolution on beamline NE3A at PF-AR from a crystal obtained in Index condition No. 5 after three weeks incubation. The structure was solved by the molecular-replacement

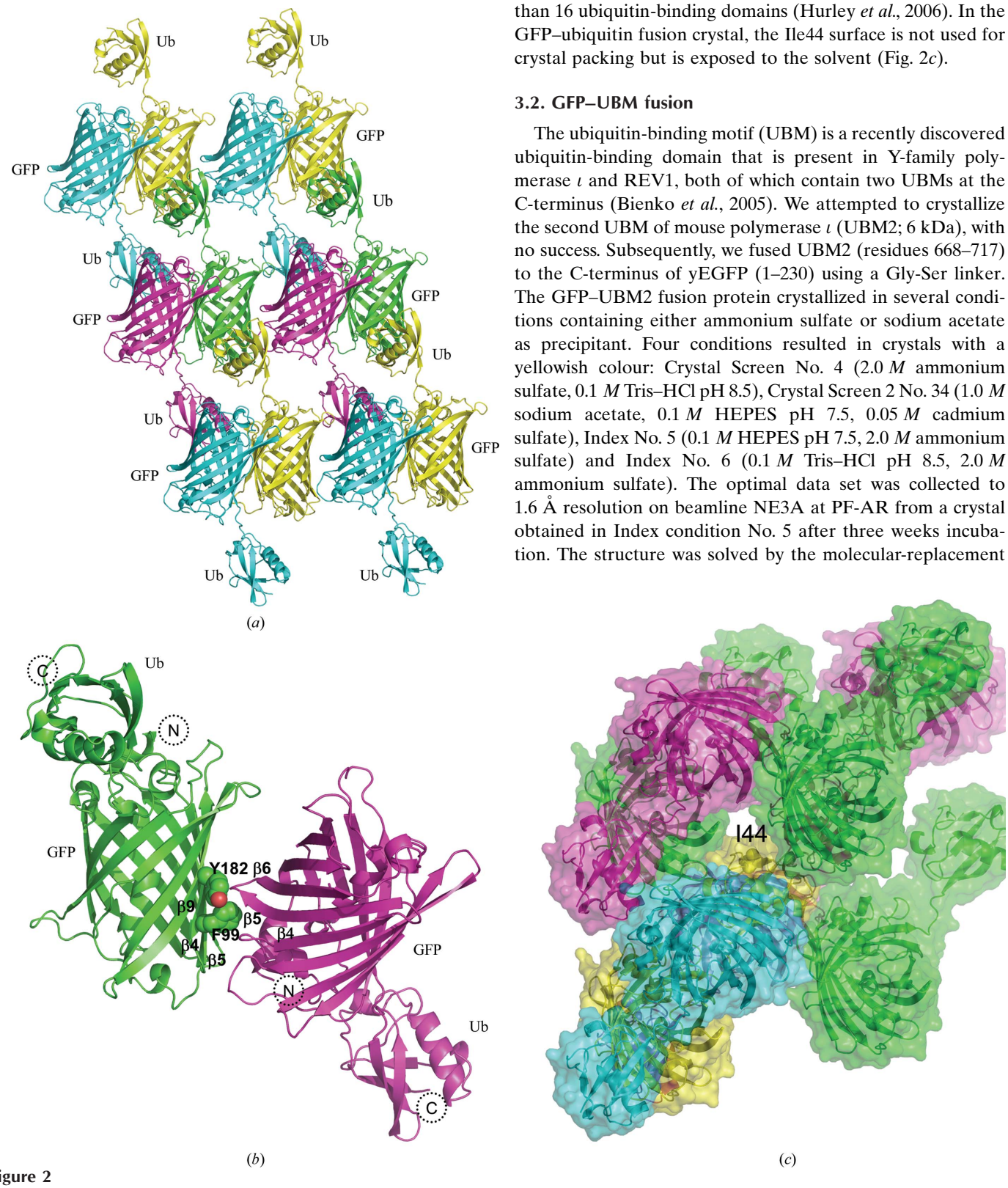

Figure 2
(*a*) Crystal packing of GFP–ubiquitin fusion protein. Pairs of fusion proteins forming a dimer are coloured magenta/green or cyan/yellow. (*b*) GFP dimer–dimer interface of GFP–ubiquitin fusion proteins. The side chains of Phe99 and Tyr182 involved in the interaction between two GFP dimers are shown as sphere models. The N- and C-termini of the fusion protein are indicated by dotted circles. (*c*) Molecular surfaces of GFP–ubiquitin molecules surrounding the central molecule (yellow). Each molecule is colour-coded as in (*a*). The Ile44 residue of ubiquitin is represented as a sphere model.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	GFP–ubiquitin	GFP–UBM2
Data collection		
No. of crystals	1	1
Beamline	PF-AR NW12A	PF-AR NE3A
Wavelength (Å)	1.0000	1.0000
Detector	ADSC Q210	ADSC Q270
Crystal-to-detector distance (mm)	117.10	133.90
Rotation range per image (°)	1	0.5
Total rotation range (°)	180.0	112.5
Exposure time per image (s)	3	2
Resolution (Å)	100–1.40 (1.42–1.40)	100–1.60 (1.63–1.60)
Space group	$C2$	$P4_22_12$
Unit-cell parameters (Å, °)	a = 92.5, b = 65.6, c = 70.2, β = 130.3	$a = b$ = 87.5, c = 83.6
Mosaicity (°)	0.375	0.550
Total No. of reflections	231179	343318
Unique reflections	62484	39422
Redundancy	3.7 (3.6)	8.7 (8.9)
Mean $I/\sigma(I)$	22.3 (2.6)	29.0 (6.3)
Completeness (%)	99.1 (100.0)	90.7 (100.0)†
$R_{\rm merge}$‡ (%)	5.8 (54.5)	6.3 (40.3)
$R_{\rm r.i.m.}$§ (%)	6.8 (64.0)	6.7 (42.7)
Overall B factor from Wilson plot (Å^2)	16.5	18.7
Refinement		
R¶ (%)	17.9	18.9
$R_{\rm free}$†† (%)	20.9	22.1
No. of atoms		
Total	2708	2372
Protein	2389	2071
Ligand	20 (ethylene glycol)	5 (SO_4^{2-})
Water molecules	299	296
Average B factors (Å^2)		
Overall	15.9	16.9
Protein	14.4	15.5
Ligand	29.5	44.5
Water molecules	26.9	26.9
R.m.s.d.		
Bond lengths (Å)	0.019	0.018
Bond angles (°)	1.88	1.86

† The relatively low overall completeness of the GFP–ubiquitin data set was caused by the presence of ice rings. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement. § $R_{\rm r.i.m.}$ (the redundancy-independent $R_{\rm merge}$) $= \sum_{hkl}[N/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ¶ $R = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$. †† $R_{\rm free}$ was calculated using a randomly selected 5% of reflections that were excluded from refinement.

method using GFP as a search model followed by automatic model building using *ARP*/*wARP* (Perrakis *et al.*, 1999), which successfully modelled 25 of the 29 residues of UBM2 (Table 1).

The Gly-Ser linker between GFP and UBM2, the ten N-terminal residues of UBM2 (residues 668–677) and the 11 C-terminal residues of UBM2 (707–717) are not resolved in this structure. The GFP moieties of the GFP–UBM2 fusion also form a dimer, as observed in the GFP–ubiquitin structure (Fig. 3*a*). The UBM2 domain (678–706) folds into a compact shape consisting of an N-terminal loop and two helices that are arranged in a V shape with the N-terminal loop in between them (Fig. 3*b*).

The N-terminal loop and α2 helix of UBM2 interact with the GFP β-barrel surface *via* residues Thr97, Phe99 (in the β4 strand) and Tyr182 (in the β9 strand) (Fig. 4*b*). Because the 12 residues between GFP and UBM2 are not resolved in the crystal, the intramolecular or intermolecular character of this interaction cannot be clearly established. UBM2 forms a dimer in the GFP–UBM2 fusion crystal: the two V-shaped helices of UBM2 interact with the counterpart molecule in an upside-down manner, significantly contributing to the crystal lattice formation (Figs. 3*a* and 3*b*). Overall, 29% (773 of 2631 Å^2) of the UBM2 surface is engaged in interactions with GFP and 13% (353 Å^2) of the surface aids UBM2 dimerization. This contrasts with the GFP–ubiquitin crystal, in which only 20% of the ubiquitin surface is engaged in crystal contacts. As in the case of GFP–ubiquitin, GFP helps in the crystal lattice formation of GFP–UBM2 by GFP dimerization (820 Å^2) and the formation of other GFP–GFP contacts (468 Å^2).

Recently, the solution structures of human UBM2 alone and in complex with ubiquitin have been reported (Bomar *et al.*, 2010). The overall structure of mouse UBM2 in our GFP–UBM2 fusion crystal superposes well with the solution structure of human UBM2, with the exception of the N-terminal loop (Fig. 4*a*). Although the different conformations in this region could be accounted for by sequence differences

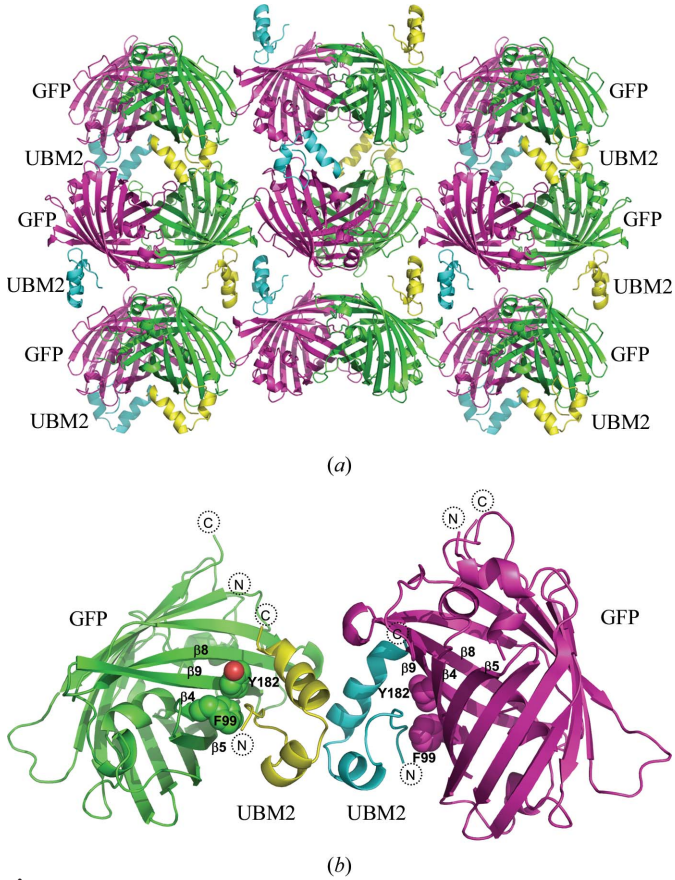

Figure 3
(*a*) Crystal packing of the GFP–UBM2 fusion protein. Pairs of GFP moieties forming a dimer are coloured magenta and green. The UBM2 moieties interacting with the respective GFP moieties are coloured cyan and yellow. (*b*) Dimerization of UBM2 moieties in the GFP–UBM2 crystal. GFP and UBM2 moieties are coloured green/magenta and yellow/cyan, respectively; their N- and C-termini are marked by dotted circles. The side chains of Phe99 and Tyr182 of GFP involved in the interaction with UBM2 are shown as sphere models.

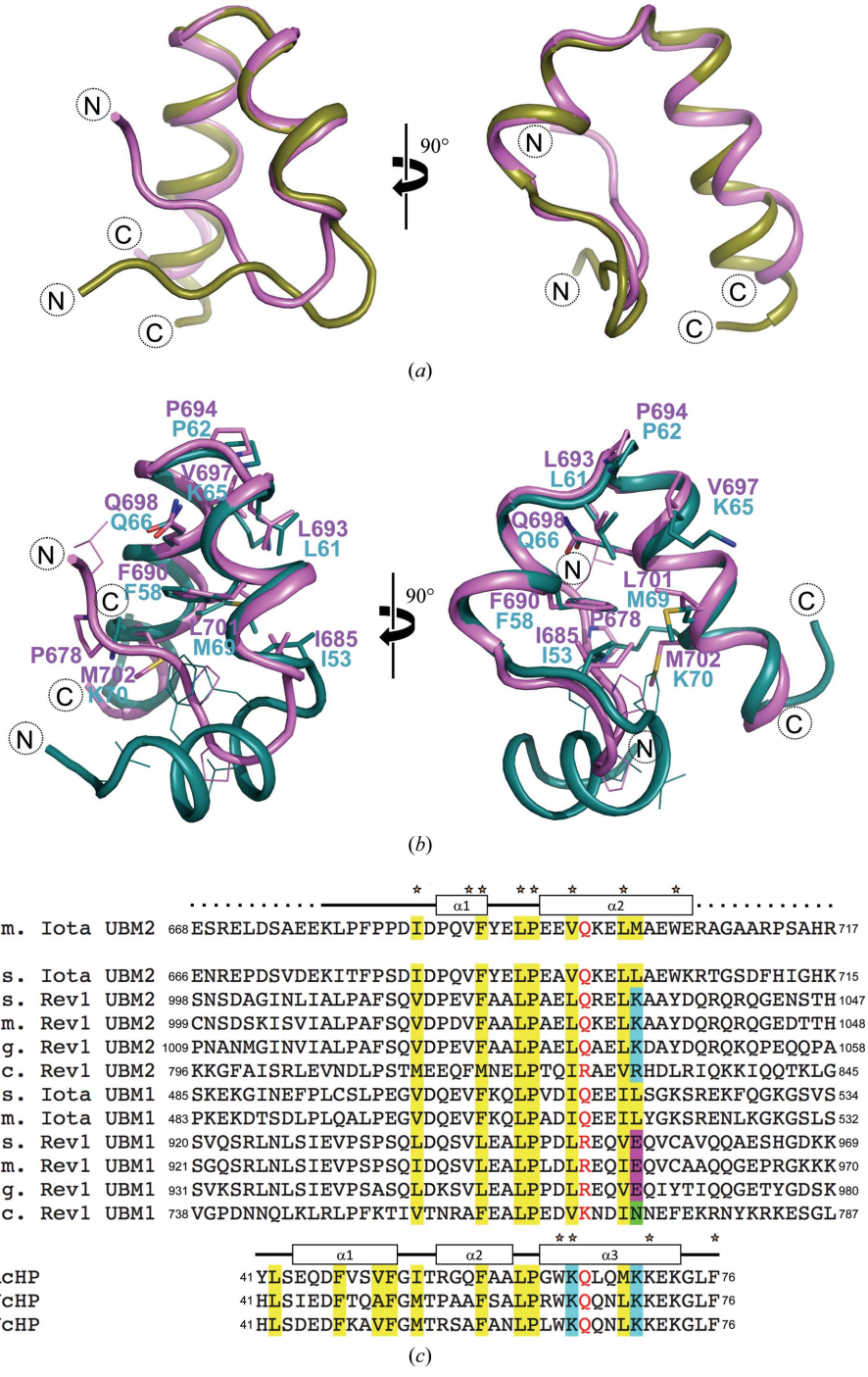

Figure 4
(*a*) Structural comparison between mouse polymerase ι UBM2 (magenta) determined using X-ray diffraction data and human polymerase ι UBM2 (dark yellow) determined using NMR data. The overall structures superpose well, with an r.m.s.d. on C^{α}-atom positions of 0.50 Å, implying that crystal packing has little effect on the overall structure of mouse polymerase ι UBM2. The N-terminal regions that show differences in conformation have different amino-acid sequences, as shown in (*c*). (*b*) Superposition of UBM2 (magenta) and HAcHP (dark cyan). The side chains contributing to the formation of the hydrophobic core in HAcHP and the corresponding residues and Pro678 in UBM2 are shown as stick models. (*c*) Amino-acid sequence alignment of UBMs from *Homo sapiens* (H.s.), *Mus musculus* (M.m.), *Gallus gallus* (G.g.) and *Saccharomyces cerevisiae* (S.c.) and C-terminal headpiece subdomains from human advillin (HAcHP), human villin (HVcHP) and chicken villin (CVcHP). Hydrophobic and basic residues involved in the formation of the hydrophobic core in HAcHP are highlighted in yellow and cyan, respectively. The corresponding residues in α1 and α2 in UBM2 are also highlighted in yellow (hydrophobic), cyan (basic) and green (polar). Gln66, which is wedged between α2 and α3 in cHPs, and the corresponding residues in UBMs are shown in red. The target protein binding sites are labelled with stars.

between the human (TFPS) and mouse (PFPP) proteins (Fig. 4*c*), we suspect that the conformation of the loop in mouse UBM2 is distorted by the interaction with GFP. The solution structure of UBM2 in complex with ubiquitin further demonstrated that the dimerization interface between the UBM2 moieties in the GFP–UBM2 fusion crystal (Fig. 3*b*) is used for ubiquitin binding.

A structural homology search using the *DALI* algorithm (Holm *et al.*, 2008) showed that UBM2 is remarkably similar (Z = 2.0, r.m.s.d. = 1.5 Å) to the C-terminal headpiece subdomain of human advillin (HAcHP; PDB code 1und; Vermeulen *et al.*, 2004). The structures of the two helices in these proteins (α1 and α2 in UBM2 and α2 and α3 in HAcHP) superpose well. While the N-terminus of UBM2 makes an extended loop, the same region in HAcHP forms an α-helix (Figs. 3*b* and 3*c*). Interestingly, HAcHPs is considered to bind F-actin using the same surface that UBMs use for ubiquitin binding (Fig. 4*c*).

3.3. Comparison with yellow fluorescent protein–glutaredoxin fusion

To date, only two structures of GFP (and its derivative) fusions have been deposited in the PDB. These examples are fusions of yellow fluorescent protein (YFP) with glutaredoxin (PDB code 2jad; Håkansson & Winther, 2007) and the GCaMP2 calcium sensor (Wang *et al.*, 2008; Akerboom *et al.*, 2009; see §4). The structure of the YFP–glutaredoxin fusion protein does not contain any intramolecular interactions between the YFP (27 kDa) and the glutaredoxin (12 kDa). The two proteins are connected by an eight-residue linker (four Ser-Gly repeats; Håkansson *et al.*, 2006). This lack of intramolecular interactions is similar to that observed in the GFP–ubiquitin fusion crystal. Crystal lattice formation is predominantly mediated by intermolecular interactions between one YFP–glutaredoxin fusion and the neighbouring YFP moiety. The N-terminal α-helix preceding strand β1 and the short

helices before strand β4 of one YFP moiety interact extensively with the β7 and β10 strands of the second YFP moiety (Fig. 5). Noticeably, Tyr200 in the β10 strand of the second YFP is centred in the interface and Tyr151 in the β7 strand periphery is also involved in the interaction (Fig. 5). In contrast, the glutaredoxin moiety of the first molecule interacts with the β2, β3 and β11 strands of the second YFP. Here, the side chain of Tyr26 of glutaredoxin packs against the β-barrel surface of YFP (Fig. 5).

In the structure of the GCaMP2 calcium sensor, in which a circularly permutated enhanced GFP is fused with a calmodulin-binding peptide and calmodulin at the N-terminus and the C-terminus, respectively, the GCaMP2 showed several different crystal forms depending on whether it was in a monomeric or a dimeric state (Wang *et al.*, 2008; Akerboom *et al.*, 2009). The oligomerization is influenced by Ca^{2+} conditions and mutations introduced into surface residues; even so, both of the aromatic GFP residues Phe99 and Tyr182 are involved in crystal contacts. Taken together, regardless of the different types of crystal packing, aromatic residues seem to play important roles in all fusion-protein structures.

3.4. The tyrosine belt on the GFP surface is a potential protein–protein interaction site

There are six aromatic residues on the GFP surface whose side chains are orientated away from the molecule: Tyr39 in the loop between the β2 and β3 strands, Phe99 in strand β4, Tyr151 in strand β7, Tyr182 in strand β9, Tyr200 in strand β10 and Phe223 in strand β11. Frequently, GFP forms an antiparallel dimer in the crystal using the β10 and β11 strands and the loop before strand β7; in this arrangement, Tyr39 and Phe223 are observed in and around the interaction interface (Fig. 6). Therefore, these two aromatic residues are usually engaged in the dimerization of GFP. In the case of the YFP–glutaredoxin fusion crystal, Tyr39 and Phe223 from the neighbouring YFP interact with the C-terminal residues (His231–Lys238) of YFP (Fig. 5). These observations suggest that the Tyr residues on the GFP surface could act as potential interaction sites for other proteins. Interestingly, the side chains of the four aromatic residues that are not involved in dimerization (Phe99, Tyr182, Tyr151 and Tyr200) are aligned on the β-barrel surface of GFP, with only one polar residue (Asn164) in the middle of this arrangement (Fig. 6). We refer to this array of aromatic residues as the 'tyrosine belt' and suggest that it could serve as an interaction platform for a variety of other proteins.

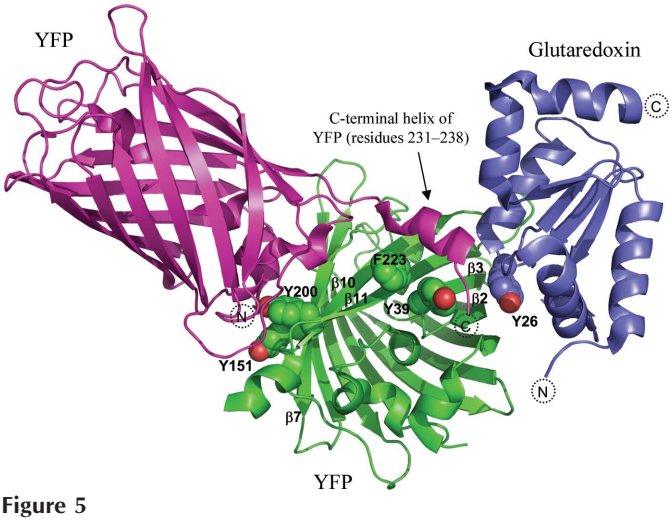

Figure 5
Crystal packing of the YFP–glutaredoxin fusion protein (Håkansson & Winther, 2007). The YFP and glutaredoxin moieties are coloured magenta and violet, respectively. The neighbouring second YFP moiety is shown in green. The side chains of Tyr39, Tyr151, Tyr200 and Phe223 of the second YFP involved in interaction with the first YFP are shown as sphere models. The side chain of glutaredoxin Tyr26 interacting with the second YFP is also shown as a sphere model.

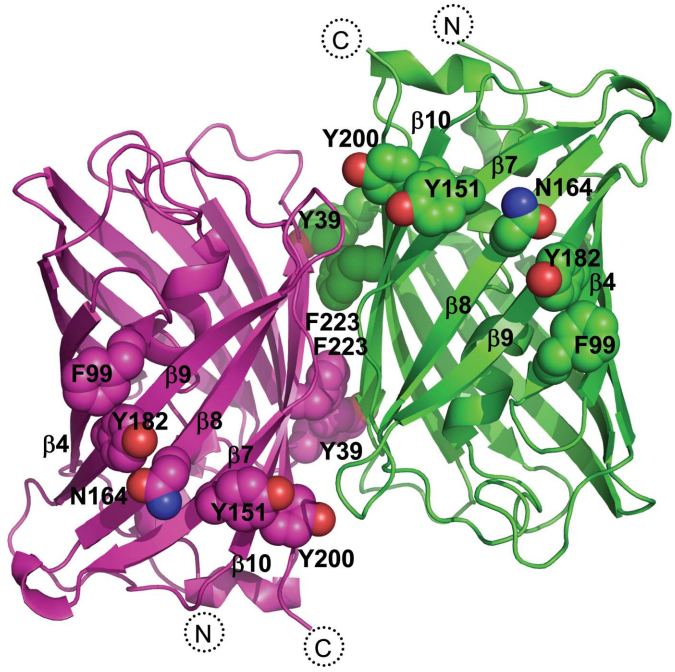

Figure 6
Tyrosine belt on the surface of GFP. The two GFP molecules in a dimer are coloured green and magenta. Arrays of Phe99, Tyr182, Tyr151 and Tyr200 side chains with Asn164 in the middle are shown as sphere models. The side chains of Ty39 and Phe223 involved in GFP dimerization are also shown.

4. Discussion

Here, we have demonstrated that GFP can be successfully used as a tag for the crystallization of small proteins or domains. Crystal lattice formation of the GFP-fusion protein is assisted by the Tyr belt on the β-barrel surface of GFP. Tyrosine has previously been successfully introduced to create artificial protein–protein interactions (Koide, 2009) and Fellouse and coworkers have demonstrated that a binary code consisting of only tyrosine and serine residues can produce high-affinity antibodies that recognize target proteins (Fellouse *et al.*, 2005). Structural analyses have also revealed that the large side chain of tyrosine contributes surprisingly well to molecular recognition (Fellouse *et al.*, 2005). This Tyr/Ser binary code has been successfully transplanted into a non-

antibody scaffold protein, termed the 'Y/S monobody', demonstrating the general effectiveness of the Tyr/Ser binary code (Koide *et al.*, 2007).

In the examples presented here, the Tyr residues in the Tyr belt of GFP resulted in the formation of a versatile molecular interface that is required for the formation of the crystal lattice of GFP–ubiquitin, GFP–UBM2 and the previously reported YFP–glutaredoxin and GCaMP2 crystals. In addition to the structures of GFP-fusion proteins, the importance of the Tyr belt has been noted in antibody derivatives: the so-called nanobodies that bind GFP and affect its fluorescence (Kirchhofer *et al.*, 2010). One of these nanobodies, which reduces GFP fluorescence by regulating the protonation state of the chromophore, binds to Tyr151 and Tyr200 in the Tyr belt, providing another example of protein–protein interactions induced through this structural feature. GFP may therefore become a useful tool in macromolecular crystallography, enabling the structural determination of proteins that are otherwise recalcitrant to crystallization.

We gratefully thank Leonard M. G. Chavas and Simon Miller for critical reading of the manuscript and valuable suggestions. We wish to thank the staff of the National Synchrotron Radiation Research Center, Taiwan and the Photon Factory, Japan for use of the macromolecular crystallographic data-collection facilities. This work was supported in part by the Strategic International Cooperative Program of JST, Grants-in-Aid for Scientific Research and Targeted Proteins Research Program from MEXT.

References

Akerboom, J., Rivera, J. D., Guilbe, M. M., Malavé, E. C., Hernandez, H. H., Tian, L., Hires, S. A., Marvin, J. S., Looger, L. L. & Schreiter, E. R. (2009). *J. Biol. Chem.* **284**, 6455–6464.

Bienko, M., Green, C. M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kannouche, P., Wider, G., Peter, M., Lehmann, A. R., Hofmann, K. & Dikic, I. (2005). *Science*, **310**, 1821–1824.

Bomar, M. G., D'Souza, S., Bienko, M., Dikic, I., Walker, G. C. & Zhou, P. (2010). *Mol. Cell*, **37**, 408–417.

Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). *Science*, **318**, 1258–1265.

Cormack, B. P., Bertram, G., Egerton, M., Gow, N. A., Falkow, S. & Brown, A. J. (1997). *Microbiology*, **143**, 303–311.

Derewenda, Z. S. (2010). *Acta Cryst.* D**66**, 604–615.

Donahue, J. P., Patel, H., Anderson, W. F. & Hawiger, J. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 12178–12182.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Falini, G., Fermani, S., Tosi, G., Arnesano, F. & Natile, G. (2008). *Chem. Commun.*, pp. 5960–5962.

Fellouse, F. A., Li, B., Compaan, D. M., Peden, A. A., Hymowitz, S. G. & Sidhu, S. S. (2005). *J. Mol. Biol.* **348**, 1153–1162.

Håkansson, K. O., Østergaard, H. & Winther, J. R. (2006). *Acta Cryst.* F**62**, 920–922.

Håkansson, K. O. & Winther, J. R. (2007). *Acta Cryst.* D**63**, 288–294.

Hiraki, M. *et al.* (2006). *Acta Cryst.* D**62**, 1058–1065.

Holm, L., Kääriäinen, S., Rosenström, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.

Hurley, J. H., Lee, S. & Prag, G. (2006). *Biochem. J.* **399**, 361–372.

Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., Ijzerman, A. P. & Stevens, R. C. (2008). *Science*, **322**, 1211–1217.

Kawate, T. & Gouaux, E. (2006). *Structure*, **14**, 673–681.

Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C. S., Cardoso, M. C., Leonhardt, H., Hopfner, K. P. & Rothbauer, U. (2010). *Nature Struct. Mol. Biol.* **17**, 133–138.

Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. & Koide, S. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 6632–6637.

Koide, S. (2009). *Curr. Opin. Struct. Biol.* **19**, 449–457.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Moon, A. F., Mueller, G. A., Zhong, X. & Pedersen, L. C. (2010). *Protein Sci.* **19**, 901–913.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Newstead, S., Kim, H., von Heijne, G., Iwata, S. & Drew, D. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 13936–13941.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H.-J., Yao, X.-J., Weis, W. I., Stevens, R. C. & Kobilka, B. K. (2007). *Science*, **318**, 1266–1273.

Sennhauser, G. & Grütter, M. G. (2008). *Structure*, **16**, 1443–1453.

Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. & Kobe, B. (2003). *Protein Sci.* **12**, 1313–1322.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Vermeulen, W., Vanhaesebrouck, P., Van Troys, M., Verschueren, M., Fant, F., Goethals, M., Ampe, C., Martins, J. C. & Borremans, F. A. (2004). *Protein Sci.* **13**, 1276–1287.

Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). *J. Mol. Biol.* **194**, 531–544.

Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. (1999). *Nature Biotechnol.* **17**, 691–695.

Wang, Q., Shui, B., Kotlikoff, M. I. & Sondermann, H. (2008). *Structure*, **16**, 1817–1827.

Yang, F., Moss, L. G. & Phillips, G. N. Jr (1996). *Nature Biotechnol.* **14**, 1246–1251.