Cloning, expression, crystallization and preliminary X-ray studies of a superfolder GFP fusion of cyanobacterial Psb32

Pasqual Liauw,^a Daniela Kannchen,^a Raphael Gasper,^b Nina Dyczmons-Nowaczyk,^a Marc M. Nowaczyk^a and Eckhard Hofmann^{b*}

Received 23 December 2014

Accepted 25 February 2015

Edited by T. Bergfors, Uppsala University, Sweden

Keywords: Psb32; superfolder GFP; *Thermosynechococcus elongatus* BP-1.

Supporting information: this article has supporting information at journals.iucr.org/f

^aDepartment of Plant Biochemistry, Ruhr-University Bochum, 44780 Bochum, Germany, and ^bDepartment of Biophysics, Ruhr-University Bochum, 44780 Bochum, Germany. *Correspondence e-mail: eckhard.hofmann@bph.rub.de

A fusion of Psb32 from the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1 (*TePsb32*) with superfolder GFP was created for enhanced solubility and improved detection and purification. The fusion protein readily formed large hexagonal crystals belonging to space group $P6_122$. A full data set extending to 2.3 Å resolution was collected at the Swiss Light Source. The phase problem could be solved by using only the sfGFP fusion partner or by using GFP and *AtTLP18.3* from *Arabidopsis thaliana* as search models. Based on this expression construct, a versatile library of 24 vectors combining four different superfolder GFP variants and three affinity tags was generated to facilitate expression and screening of fluorescent fusion proteins.

1. Introduction

Generation of fusion proteins has become a routine approach in many areas of molecular and cell biology. In particular, fusion of fluorescent proteins is frequently used for protein localization in the cellular context or determination of protein interactions. This approach has also proven to be beneficial in structural studies to speed up the process of protein detection (Chalfie *et al.*, 1994; Drew *et al.*, 2006; Kawate & Gouaux, 2006; Bird *et al.*, 2015) and even crystallization and phasing (Håkansson & Winther, 2007; Suzuki *et al.*, 2010; Mueller *et al.*, 2013). Especially in the case of fusion proteins, the design of proper expression vectors (containing target protein, affinity tags and fusion partner) is crucial for both expression and purification.

We therefore designed a highly customizable modular vector system based on the pRSET T7 family of expression vectors (Schoepfer, 1993). The combination of a standard T7 expression system with a modular plasmid architecture offers high flexibility and high expression yields. Whilst expression of the protein of interest (POI) with an affinity tag on either the N- or C-terminus is a standard technique, here we combine these tags with the solubility enhancer superfolder GFP (sfGFP; Pédelacq *et al.*, 2006), which can boost the expression of soluble target proteins. This has been shown for *Tobacco etch virus* (TEV) protease (Wu *et al.*, 2009), for example. Owing to the ease of GFP quantification by fluorescence measurements (Coutard *et al.*, 2008), protein concentration can readily be determined for such fusion proteins and can speed up the optimization of culture conditions. Furthermore, a sophisticated combination of different fluorescent proteins can be used for protein–protein interaction analysis by FRET (Pollok & Heim, 1999). In addition, the (sf)GFP moiety might

also facilitate structure determination by X-ray crystallography (Suzuki *et al.*, 2010). For cases in which the large fusion tag is not compatible with downstream analyses, the TEV protease recognition site is placed between the POI and the fusion tag.

The membrane-anchored protein *TePsb32* (tl0404) from the thermophilic cyanobacterium *Thermosynechococcus elongatus* has been assigned to a class of auxiliary proteins involved in the repair of damaged PSII subunits and PSII reassembly, as has been described elsewhere (Nixon *et al.*, 2010; Nickelsen & Rengstl, 2013). This assignment is based on its homology to *SynPsb32* (sll1390) from the cyanobacterium *Synechocystis* sp. PCC 6803 and the thylakoid lumen protein *AtTLP18.3* from *Arabidopsis thaliana* (Sirpiö *et al.*, 2007; Wegener *et al.*, 2011; sequence identities of 42 and 32%, respectively). These proteins share a so-called TPM domain (Pfam PF04536) with unclear function. For *AtTLP18.3* an acid phosphatase function has been proposed (Wu *et al.*, 2011), but its physiological role is not confirmed. Specifically for cyanobacteria, evidence for (de)phosphorylation events occurring in the thylakoid lumen is elusive (Pursiheimo *et al.*, 1998; Spetea & Lundin, 2012). We therefore set out to determine the three-dimensional structure of *TePsb32* to verify its assignment as a TPM domain protein and to gain insights into structural differences from the plant homologue.

In the present study, we expressed *TePsb32* as a *Strep*-tag II–superfolder GFP fusion protein (STII-sfGFP-*TePsb32*) in

Escherichia coli and report its successful crystallization and phasing.

2. Materials and methods

2.1. Construction of expression vectors

All plasmids created in this study are based on the pRSET6a expression vector (Schoepfer, 1993) that allows gene expression under the control of an IPTG-inducible T7 promoter. For the generation of expression plasmids with an N-terminal affinity tag, the coding sequence for the tag (6×His tag, *Strep*-tag II or Twin-*Strep*-tag) was introduced between the NdeI and NheI restriction-enzyme sites, that for the fluorescent protein [sfG(reen)FP, sfB(lue)FP, sfY(ellow)FP and sfC(yan)FP] between the NheI and EcoRI sites and that for the linker and the TEV protease recognition site (Parks *et al.*, 1994; Kapust *et al.*, 2002) between the EcoRI and SfoI (EheI) sites. The coding sequence of *TePsb32* was cloned between the SfoI and KpnI sites (Table 1). Plasmids for C-terminal tagging were generated by introducing the coding sequence for the affinity tag between the SfoI and ApaI sites and that for the POI (including a GSAM linker; Corsini *et al.*, 2008) between the NdeI and NheI sites based on the plasmid for N-terminal cloning. The modular concept for the different expression vectors is illustrated in Fig. 1. The coding sequence for sfGFP (Pédrelacq *et al.*, 2006) including an N-terminal

Figure 1

Schematic representation of the modular architecture of the presented plasmids. Plasmid variants were engineered for the expression of the protein of interest (POI) with different affinity tags (6×His-tag, *Strep*-tag II or Twin-*Strep*-tag) and superfolder GFP colour variants at (a) the amino-terminus or (b) the carboxy-terminus. Plasmid pStrepGFP-*TePsb32* used for expression of STII-sfGFP-*TePsb32* (with *TePsb32* being the POI) is highlighted by a red dashed line. The 'X' in the plasmid name serves as an indicator of the respective colour variant: G, green; Y, yellow; B, blue; C, cyan.

Table 1

Macromolecule-production information.

Source organism	<i>T. elongatus</i> BP-1
DNA source	Genomic DNA
Forward primer†	GGCGCCACCACTGCAATCGATATCCC
Reverse primer‡	GGTACCTTAGGAGCGATCGTCCGTTT
Cloning vector	pJET1.2/blunt cloning vector
Expression vector	pStrepGFP- <i>TePsb32</i>
Expression host	<i>E. coli</i> C43 (DE3)
Complete amino-acid sequence of the construct produced§	MASWSHPQFEK ASMSKGEELFTGVVPILVELDGD- VNGHKFSVRGEGEGDATNGKLTGKLTCTTGKL- PVPWPTLVTTLTGVCFSRYPDHMKRHDFFK- SAMPEGYVQERTISFKDDGTGKTRAEVKFEGD- TLVNRIELKGIDFKEDGNILGHKLEYNFSHN- VYITADKQKNGIKANFKIRHNVEDGSVQLADH- YQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNE- KRDHMLLEFVTAAGITHGMDLEYKYRIRSGG- GGenlyfqaTSAIDIPFGTATGVIDEGNVL- SAVTQGSVGRSLQDLSEATGINVHVVTLRDL- YGETPQSFVDDLFSQWFPDPESQANQVIALD- TVTNGTAIHYGDAVAERLNPETAESIVQETMR- VPLREGNYNQAVLDTVDRLGKVLKGEPPDGP- VVREVVVEKTYKSKEETDDRS

† The SfoI site is underlined. ‡ The KpnI site is underlined. § STII is shown in bold, sfGFP in italics and the TEV site in lower case; *TePsb32* is underlined.

Table 2

Crystallization.

Method	Hanging-drop vapour diffusion
Plate type	Greiner 24-Well ComboPlate
Temperature (K)	291
Protein concentration (mg ml ⁻¹)	25
Buffer composition of protein solution	50 mM MES pH 6.5
Composition of reservoir solution	0.75 M sodium citrate, 0.1 M CHES pH 9.5
Volume and ratio of drop	2 µl (1:1)
Volume of reservoir (µl)	750

Twin-Strep-tag was synthesized as a dsDNA string (Life Technologies) and subcloned into the pJET1.2/blunt cloning vector (CloneJET PCR Cloning Kit, Thermo Scientific). The coding sequences for the N- and C-terminal affinity tags (6×His and *Strep*-tag II) were exchanged by PCR using the appropriate oligonucleotide primers (see Supplementary Table S1). The sfBFP (Y66H) and sfCFP (Y66W) variants were generated by primer-directed mutagenesis with the oligonucleotide primer pairs *bfp_for/bfp_rev* and *cfp_for/cfp_rev* (Pédalacq *et al.*, 2006), whereas sfYFP (T203Y) was synthesized as a dsDNA string (Life Technologies) carrying a C-terminal Twin-Strep-tag. This system allows easy exchange of N- and C-terminal affinity tags and fluorescent protein fusions.

2.2. Protein expression and purification

E. coli C43 (Lucigen Inc., USA) cells were transformed with plasmid pStrepGFP-*TePsb32* and cultivated on lysogeny broth (LB) agar plates containing 100 µg ml⁻¹ ampicillin at 310 K. For large-scale expression, colonies of an LB agar plate were resuspended in 5 ml LB medium and used for inoculation of a 100 ml pre-culture. After 3 h of constant shaking at 310 K, the main culture was inoculated [1% (v/v)] with the preculture. When the OD₆₀₀ had reached a value of 0.6, expression of

STII-sfGFP-*TePsb32* was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside. The cells were allowed to grow for 18 h under constant shaking at 303 K. Following centrifugation, the cell pellet was washed with 100 mM Tris-HCl pH 8.0 supplemented with 500 mM sucrose, centrifuged again and subsequently resuspended in 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride. After cell disruption by sonication, cell debris and insoluble particles were removed by centrifugation followed by sterile filtration (0.45 µm membrane pore size). The lysate was applied onto a 5 ml Strep-Tactin Superflow high-capacity cartridge (IBA Lifesciences) using a GE Healthcare ÄKTApurifier 100 at room temperature. After washing the column with ten column volumes of buffer *A* (100 mM Tris-HCl pH 8.0, 150 mM NaCl), bound proteins were eluted using buffer *B* (buffer *A* + 2.5 mM D-desthiobiotin; IBA Lifesciences). Pooled fractions were dialyzed two times against 1 l 50 mM MES pH 6.5 and the sample was subsequently applied onto a RESOURCE Q column (GE). Bound protein was eluted with a linear gradient from 0 to 1 M NaCl in 50 mM MES pH 6.5. The highly pure protein with a size of 48.6 kDa (as visible on SDS-PAGE; Fig. 2) was concentrated to 10 mg ml⁻¹ and used for initial crystallization screening. Following the abovementioned protocol, roughly 50 mg pure fusion protein could be obtained per litre of expression culture.

2.3. Electrophoresis

Proteins were separated on 15% SDS polyacrylamide gels according to Laemmli (1970). In order to retain sfGFP fluorescence during electrophoresis, the samples were not heated and were applied to the gel in 10% (w/v) glycerol. The fusion proteins were visualized either by staining with 0.25% Coomassie Brilliant Blue (CBB) R-250 or by irradiation with

Figure 2

Native STII-sfGFP-*TePsb32* appeared as a single protein of around 48 kDa on 15% SDS-PAGE after elution from a RESOURCE Q column. Fermentas PageRuler Prestained Protein Ladder was used as a marker.

Table 3

Data collection and processing.

Values in parentheses are for the highest resolution shell.

Diffraction source	X10SA, SLS
Wavelength (Å)	0.95107
Temperature (K)	100
Detector	PILATUS 6MF
Crystal-to-detector distance (mm)	435
Rotation range per image (°)	0.25
Total rotation range (°)	360
Exposure time per image (s)	0.1
Space group	<i>P</i> 6 ₁ 22
<i>a</i> , <i>b</i> , <i>c</i> (Å)	144.74, 144.74, 91.35
α , β , γ (°)	90, 90, 120
Mosaicity (°)	0.113
Resolution range (Å)	47.4–2.3 (2.36–2.30)
Total No. of reflections	980769 (70376)
No. of unique reflections	25440 (1822)
Completeness (%)	99.5 (99.1)
Multiplicity	38.6 (38.6)
$\langle I/\sigma(I) \rangle$	28.1 (2.2)
R_{meas} (%)	13.0 (312.7)
$CC_{1/2}$	100 (75)
Overall <i>B</i> factor from Wilson plot (Å ²)	63.6

UV light at 365 nm in an Alpha Innotech ChemiImager Ready imaging system.

2.4. Crystallization

Crystallization screening was carried out by the sitting-drop vapour-diffusion method using a Phoenix crystallization robot [Art Robbins Instruments, Sunnyvale, USA; 0.1 µl protein was mixed with 0.1 µl crystallization solution (Qiagen) at 291 K].

Figure 3

Hexagonal rod-shaped crystal of STII-sfGFP-*TePsb32* (400 × 50 × 45 µm). Crystallization was performed by hanging-drop vapour diffusion in 0.75 *M* sodium citrate, 0.1 *M* CHES pH 9.5 at 291 K using a protein concentration of 25 mg ml^{−1}.

Small crystals were observed in The JCSG Core Suite II (1 *M* sodium citrate, 0.1 *M* CHES pH 9.5) after 2 d. After optimization of the crystallization conditions, hexagonal rod-shaped crystals (400 × 50 × 45 µm; Fig. 3) were grown in 0.75 *M* sodium citrate, 0.1 *M* CHES pH 9.5 using a protein concentration of 25 mg ml^{−1} by hanging-drop vapour diffusion at 291 K (Table 2). The integrity of the content of the protein crystals was confirmed by SDS–PAGE analysis (Fig. 4).

2.5. X-ray data collection

For data collection under cryogenic conditions, crystals were briefly soaked in mother liquor containing increasing concentrations of glucose (15–30%). A data set was collected from a single crystal on beamline X10SA at the Swiss Light Source (SLS), Villigen, Switzerland. Data were processed with *XDS* (Kabsch, 2010) and scaled using *XSCALE* (Kabsch, 2010). Data-collection statistics are shown in Table 3.

3. Results and discussion

Protein yields from expression of *TePsb32* in *E. coli* could be increased significantly by using the newly generated plasmid pStrepGFP-*TePsb32* (Fig. 1). Whereas expression without the N-terminal sfGFP moiety yielded only 1 mg protein per litre of culture (data not shown), up to 50 mg of the STII-sfGFP-*TePsb32* fusion protein was obtained per litre of culture. In addition, optimization of expression and purification greatly benefited from the sfGFP fluorescence; for example, the efficiency of cell disruption was readily indicated by the loss of fluorescence in the pellet after centrifugation. Crystallization by hanging-drop vapour diffusion at 291 K yielded fluorescent hexagonal rod-shaped crystals (Fig. 3). A data set to a resolution of 2.3 Å was obtained from a single STII-sfGFP-*TePsb32* crystal on beamline X10SA at the Swiss Light Source, Villigen, Switzerland.

Figure 4

Analysis of dissolved crystals of STII-sfGFP-*TePsb32* on 15% SDS–PAGE demonstrates the integrity of the fusion protein. Proteins were stained with Coomassie Brilliant Blue or irradiated with UV light. Lane 1 contains the protein solution, lane 2 contains Fermentas Unstained Protein Molecular Weight Marker and lane 3 contains redissolved protein crystals.

The crystals belonged to the hexagonal space group $P6_122$ (No. 178), with unit-cell parameters $a = b = 144.74$, $c = 91.35$ Å. Based on the Matthews coefficient of $2.84 \text{ Å}^3 \text{ Da}^{-1}$, one molecule per asymmetric unit is expected, resulting in a solvent content of 56.7%.

The phase problem was solved by molecular replacement using the structure of sfGFP (Pédélecq *et al.*, 2006; PDB entry 2b3p) as a model in *Phaser* (McCoy *et al.*, 2007). Clear difference density for the missing TePsb32 domain was visible after 20 cycles of rigid-body and positional refinement of the partial solution in *PHENIX* (Afonine *et al.*, 2012). Therefore, even without *a priori* knowledge of the fold of TePsb32, a complete model of the entire fusion protein could be built.

Nevertheless, to speed up the process of structure determination, the automatic molecular-replacement pipeline *BALBES* (Long *et al.*, 2008) was used. *BALBES* combined the high-resolution structures of GFP (Shinobu *et al.*, 2010; PDB entry 2wur) and AtTLP18.3 (Wu *et al.*, 2011; PDB entry 3pvh) to provide a unique solution ($R = 0.355$, $R_{\text{free}} = 0.415$, Q factor = 0.707) containing a single STII-sfGFP-TePsb32 molecule in the asymmetric unit. Structure determination including model building and refinement using *Coot* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010) and *PHENIX* is currently in progress.

Based on the positive results for this sfGFP fusion, an extended library of 24 different expression constructs was generated, including four different superfolder GFP variants (as N- and C-terminal fusions) with three different affinity tags (Fig. 1). This library is now being tested with a number of different target proteins.

Acknowledgements

This work was supported by the Studienstiftung des deutschen Volkes (PL) and by grants from the Deutsche Forschungsgemeinschaft to MMN (Nos. 836/3-1 and 836/1-1). We thank the beamline staff of X10SA at the Swiss Light Source, Villigen, Switzerland and at the beamlines at the European Synchrotron Radiation Facility (Grenoble, France) for support during data collection. We also appreciate the excellent technical assistance by Ursula Altenfeld and Petros Sarantopoulos.

References

- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst. D* **68**, 352–367.
- Bird, L. E., Rada, H., Verma, A., Gasper, R., Birch, J., Jennions, M., Löwe, J., Moraes, I. & Owens, R. J. (2015). *J. Vis. Exp.*, e52357.
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. (1994). *Science*, **263**, 802–805.
- Corsini, L., Hothorn, M., Scheffzek, K., Sattler, M. & Stier, G. (2008). *Protein Sci.* **17**, 2070–2079.
- Coutard, B., Gagnaire, M., Guillhon, A.-A., Berro, M., Canaan, S. & Bignon, C. (2008). *Protein Expr. Purif.* **61**, 184–190.
- Drew, D., Lerch, M., Kunji, E., Slotboom, D.-J. & de Gier, J.-W. (2006). *Nature Methods*, **3**, 303–313.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst. D* **60**, 2126–2132.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486–501.
- Håkansson, K. O. & Winther, J. R. (2007). *Acta Cryst. D* **63**, 288–294.
- Kabsch, W. (2010). *Acta Cryst. D* **66**, 125–132.
- Kapust, R. B., Tözsér, J., Copeland, T. D. & Waugh, D. S. (2002). *Biochem. Biophys. Res. Commun.* **294**, 949–955.
- Kawate, T. & Gouaux, E. (2006). *Structure*, **14**, 673–681.
- Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
- Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst. D* **64**, 125–132.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Mueller, G. A., Pedersen, L. C., Lih, F. B., Glesner, J., Moon, A. F., Chapman, M. D., Tomer, K. B., London, R. E. & Pomés, A. (2013). *J. Allergy Clin. Immunol.* **132**, 1420–1426.
- Nickelsen, J. & Rengstl, B. (2013). *Annu. Rev. Plant Biol.* **64**, 609–635.
- Nixon, P. J., Michoux, F., Yu, J., Boehm, M. & Komenda, J. (2010). *Ann. Bot. (Lond.)*, **106**, 1–16.
- Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. (1994). *Anal. Biochem.* **216**, 413–417.
- Pédélecq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. (2006). *Nature Biotechnol.* **24**, 79–88.
- Pollok, B. A. & Heim, R. (1999). *Trends Cell Biol.* **9**, 57–60.
- Pursiheimo, S., Rintamäki, E., Baena-Gonzalez, E. & Aro, E.-M. (1998). *FEBS Lett.* **423**, 178–182.
- Schoepfer, R. (1993). *Gene*, **124**, 83–85.
- Shinobu, A., Palm, G. J., Schierbeek, A. J. & Agmon, N. (2010). *J. Am. Chem. Soc.* **132**, 11093–11102.
- Sirpiö, S., Allahverdiyeva, Y., Suorsa, M., Paakkarinen, V., Vainonen, J., Battchikova, N. & Aro, E.-M. (2007). *Biochem. J.* **406**, 415–425.
- Spetea, C. & Lundin, B. (2012). *FEBS Lett.* **586**, 2946–2954.
- Suzuki, N., Hiraki, M., Yamada, Y., Matsugaki, N., Igarashi, N., Kato, R., Dikic, I., Drew, D., Iwata, S., Wakatsuki, S. & Kawasaki, M. (2010). *Acta Cryst. D* **66**, 1059–1066.
- Wegener, K. M., Bennewitz, S., Oelmüller, R. & Pakrasi, H. B. (2011). *Mol. Plant*, **4**, 1052–1061.
- Wu, H.-Y., Liu, M.-S., Lin, T.-P. & Cheng, Y.-S. (2011). *Plant Physiol.* **157**, 1015–1025.
- Wu, X., Wu, D., Lu, Z., Chen, W., Hu, X. & Ding, Y. (2009). *J. Biomed. Biotechnol.* **2009**, 591923.