ACTA ABP BIOCHIMICA POLONICA

Vol. 62, No 3/2015
407–411
http://dx.doi.org/10.18388/abp.2015_1026

Regular paper

A systematic investigation of the stability of green fluorescent protein fusion proteins*

Monika Janczak[1], Michał Bukowski[1], Andrzej Górecki[2], Grzegorz Dubin[3,4], Adam Dubin[1] and Benedykt Wladyka[1,4]✉

[1]Department of Analytical Biochemistry, [2]Department of Physical Biochemistry, [3]Department of Microbiology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Kraków, Poland; [4]Malopolska Centre of Biotechnology, Jagiellonian University, Kraków, Poland

X-ray crystallography provides important insights into structure-function relationship in biomolecules. However, protein crystals are usually hard to obtain which hinders our understanding of multiple important processes. Crystallization requires large amount of protein sample, whereas recombinant proteins are often unstable or insoluble. Green fluorescent protein (GFP) fusion is one of the approaches to increase protein synthesis, solubility and stability, facilitating crystallization. In this study we analyze the influence of the linker length, composition and the position of GFP relative to the fusion partner on the fusion protein production and stability. To this end, multiple constructs of enzymatically impaired variant of $PemK_{Sa}$ toxin from *Staphylococcus aureus* CH91 fused to GFP were generated. Fusion protein production in *Escherichia coli* was evaluated. The proteins were purified and their stability tested. $PemK_{Sa}$-α14aa-GFP fusion provided best production and stability. Obtained results demonstrate the importance of optimization of fusion protein construct, including linker selection and the order of fusion partners, in obtaining high quantities of stable protein for crystallization.

Key words: crystallography, fusion protein, recombinant protein, toxin-antitoxin system

Received: 15 April, 2015; revised: 28 April, 2015; accepted: 02 May, 2015; available on-line: 21 July, 2015

INTRODUCTION

Protein crystallization followed by X-ray diffraction studies allows to elucidate the three dimensional structures, providing vital information on structure-activity relationships. Understanding the molecular details of the mechanism of action of different proteins provides important progress in basic research and multiple other fields including drug discovery (Overington *et al.*, 2006). However, crystallization requires a considerable amount of purified, well behaved, sample, which is often hard to obtain. Fusion proteins are one of the attractive tools which frequently allow to overcome this limitation (Smyth *et al.*, 2003). Fusion proteins not only facilitate purification, but are also used to improve production, solubility and stability of target proteins. Protein tags are attached either at N- or C-terminus of the fusion partners or less often replace intrinsic unstructured regions (Cherezov *et al.*, 2007). The importance of fusion proteins in protein crystallization is not limited to facilitating protein synthesis only. By increasing the available polar surface fusion proteins may facilitate protein crystals growth (Cherezov *et al.*, 2007).

Commonly used protein tags include maltose-binding protein (MBP) (Kobe *et al.*, 1999), glutathione-S-transferase (GST) (Lally *et al.*, 1998), thioredoxin A (Corsini *et al.*, 2008), antibody fragments (Iwata *et al.*, 1995) and lysozyme (Cherezov *et al.*, 2007), but other tags have also been used. These tags helped obtaining a large number of crystal structures. Yet another possibility is provided by using green fluorescent protein (GFP) as a fusion tag. GFP originates from a jellyfish *Aequorea victoria* (Cubitt *et al.*, 1995). Due to a well-defined, compact structure (Yang *et al.*, 1996) and stability in varied conditions, including broad pH range, elevated temperature and the presence of detergents (Cubitt *et al.*, 1995) it constitutes an ideal fusion partner. It is also well soluble. Though GFP has a relatively large molecular weight (~27 kDa) it usually does not interfere with the function of fused proteins (Kwolik *et al.*, 2001; Hsieh *et al.*, 2010). The above characteristics have warranted the success of GFP in studying *in vivo* tracking of proteins (Hsieh *et al.*, 2010; Ashikawa *et al.*, 2011) and a fact that GFP serves as a suitable scaffold tag in protein crystallization (Suzuki *et al.*, 2010; Mueller *et al.*, 2013; Nguyen *et al.*, 2013). The position of the tag relative to a fusion partner and the properties of a peptide linker are of considerable importance for successful production and stability (Arai *et al.*, 2001; Japrung *et al.*, 2005). Nevertheless, insufficient and partly contradictory experimental data does not support rational construct design. Some reports suggest that localization of the tag at the N-terminus of the fusion partner warrants high production and proper folding (di Guan *et al.*, 1988). Another studies argue that fusion at the C-termini of the target protein improves solubility (Japrung *et al.*, 2005). Certain data indicates that the use of flexible linker sustains functionality of the fusion partner (Robinson *et al.*, 1998), but another study demonstrated that only a rigid, alpha-helical linker provides enough spatial separation of fusion partners to facilitate independent folding (Arai *et al.*, 2001). Linker length is also of importance, but insufficient data is available to

✉e-mail: benedykt.wladyka@uj.edu.pl

*A preliminary report on the same subject was presented at the XLII Winter School "From Genome to Proteome" organized by Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, 10–14 February, 2015, Zakopane, Poland.

Abbreviations: GFP, green fluorescent protein; His-tag, polyhistidine-tag; IMAC, immobilized metal-affinity chromatography; IPTG, isopropyl β-D-1-thiogalactopyranoside; NiNTA, nickel-nitrilotriacetic acid; PCR, polymerase chain reaction; SDS-PAGE, polyacrylamide gel electrophoresis with sodium dodecyl sulfate

Table 1. Plasmid constructs used in the study.

Plasmid	Genes (in order of fusion components)	Linker	Primers	Encoded fusion protein
pETDuet-gfp-2aa-pemK$_{Sa}$	*gfpmut2* *pemK$_{Sa}$*	LQ (short)	GFP_BamHI-ForB GFP_PstI-RevB PemK_PstI-ForB PemK_HindIII-RevB	GFP-2aa-PemK$_{Sa}$
pETDuet-pemK$_{Sa}$-2aa-gfp	*pemK$_{Sa}$* *gfpmut2*	LQ (short)	PemK_BamHI-For PemK_PstI-Rev2 GFP_PstI-For GFP_HindIII-Rev2	PemK$_{Sa}$-2aa-GFP
pETDuet-gfp-α14aa-pemK$_{Sa}$	*gfpmut2* *pemK$_{Sa}$*	LQAEAAAKEAAAKA (helical)	GFP_BamHI-ForB GFP_PstI_LinH-RevB PemK_PstI-ForB PemK_HindIII-RevB	GFP-α14aa-PemK$_{Sa}$
pETDuet-pemK$_{Sa}$-α14aa-gfp	*pemK$_{Sa}$* *gfpmut2*	LQAEAAAKEAAAKA (helical)	PemK_BamHI-For PemK_PstI-Rev2 GFP_PstI-LinH-For GFP_HindIII-Rev2	PemK$_{Sa}$-α14aa-GFP
pETDuet-gfp-15aa-pemK$_{Sa}$	*gfpmut2* *pemK$_{Sa}$*	LQTSGGGGSGGGGSA (unstructured)	GFP_BamHI-ForB GFP-PstI_LinE-RevB PemK_PstI-ForB PemK_HindIII-RevB	GFP-15aa-PemK$_{Sa}$
pETDuet-pemK$_{Sa}$-15aa-gfp	*pemK$_{Sa}$* *gfpmut2*	LQTSGGGGSGGGGSA (unstructured)	PemK_BamHI-For PemK_PstI-Rev2 GFP_PstI_LinE-For GFP_HindIII-Rev2	PemK$_{Sa}$-15aa-GFP

conclude on the best length and composition (Kwolik *et al.*, 2001; Corsini *et al.*, 2008).

Clearly, the controversy concerning construct design is partly related to the fact that different fusion partners may require different strategies to obtain efficient production. Nevertheless, we believe that a systematic investigation may define certain general recommendations for rational construct design. Therefore, in this study we designed a set of fusion proteins containing GFP and an enzymatically impaired staphylococcal mRNA interferase PemK$_{Sa}$(R84A). The interferase, belonging to the toxin-antitoxin system encoded on pCH91 plasmid of *Staphylococcus aureus* CH91 (Takeuchi *et al.*, 1999; Bukowski *et al.*, 2013), was chosen due to problematic production and stability of his-tagged constructs (unpublished results). Synthesis level, solubility and stability were evaluated for N- and C-terminal GFP fusions containing different linker lengths. The results allow to formulate certain general recommendations for GFP fusion containing construct design.

MATERIALS AND METHODS

Bacterial strains and culture conditions. *E. coli* TOP10 (Invitrogen) and BL21(DE3) (Novagen) were used for cloning and fusion protein production, respectively.

Preparation of plasmid constructs. Genes encoding fusion protein partners were amplified by PCR. *pemK$_{Sa}$R84A* (following denoted as *pemK$_{Sa}$*) was amplified from pETDuet-*pemK$_{Sa}$R84A* (Bukowski *et al.*, 2013). *gfpmut2* was amplified from pCN68 (Charpentier *et al.*, 2004). Linker sequences and restriction sites were introduced within primers (Table 2). Amplicons were ligated into pTZ-57R/T (Thermo Scientific), excised with appropriate restriction enzymes and cloned into an expression vector pETDuet-1 (Novagen). The N-terminal sequence of six histidines was already encoded within the vector (Table 1 and Fig. 1).

Protein production and purification. Liquid overnight cultures were diluted 100-fold in 500 ml of Tryptic Soy Broth (Sigma Aldrich) containing ampicilin (0.1 mg/ml). The bacteria were cultured at 37°C with thorough aeration until the optical density measured at 600 nm reached 0.6. Protein production was induced with IPTG (isopropyl β-D-1-thiogalactopyranoside) at a final concentration of 1 mM. After induction, the cultures were incubated overnight at 20°C with thorough aeration. Cells were harvested by centrifugation for 30 min at 5000×*g* and lysed by sonication. Lysates were clarified for 30 min at 21 000×*g*. The fusion proteins were purified by affinity chromatography in native conditions using buffers recommended by the resin supplier (nickel-nitrilotriacetic acid, NiNTA; Novagen). The preparation was dialyzed overnight against 5 mM Tris/HCl, 50 mM NaCl, pH 7.4, precipitation was assessed, and the fusion protein was further purified by gel filtration using Superdex 75 column (GE Healthcare) in the same buffer.

Determination of stability of fusion proteins. The samples were incubated at 4°C and 22°C for two weeks. The stability and the amount of remaining fusion protein was analyzed by SDS-PAGE and densitometry. Intensity profiles were generated and quantified using ImageJ software (Schneider *et al.*, 2012).

RESULTS AND DISCUSSION

Evaluation of fusion protein production and solubility

We have previously described the toxin-antitoxin system PemIK$_{Sa}$ of *S. aureus* strain CH91 (Bukowski *et al.*, 2013), however, our attempts to obtain recombinant PemK$_{Sa}$ or its mutants at crystallization quality have

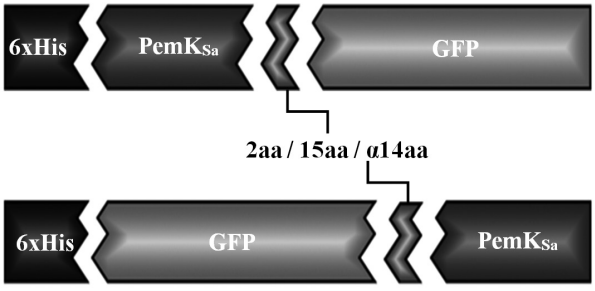

Figure 1. Schematic representation of fusion proteins evaluated in this study.
$PemK_{Sa}$ toxin and GFP were combined in two different relative orientations. In each orientation the fusion proteins were connected with three linker types (six different constructs were evaluated in total). A six histidine tag was added at the N-terminus of each fusion protein.

failed due to instability of the toxin, especially in low ionic strength buffers. In this study we ventured to obtain stable GFP fusion of $PemK_{Sa}$(R84A) (further denoted as $PemK_{Sa}$) for crystallization and in parallel to acquire a systematic evaluation of the effect of GFP position relative to the fusion protein and the linker length on production and stability. Six constructs were prepared, three containing GFP at the N-terminus of $PemK_{Sa}$ and three at the C-terminus. In both types of constructs three linkers were evaluated: a short two amino acid containing one, a flexible 15 amino acid linker and a rigid α-helical 14 amino acid containing one (Fig. 1). We analyzed the synthesis level and solubility (*vs.* inclusion bodies) of each construct in *E. coli*. The efficiency of recombinant protein production and the amount of protein in soluble fraction differed between evaluated constructs (Fig. 2 and Table 3.). Highest production was obtained for $PemK_{Sa}$-α14aa-GFP whereas production level of GFP-2aa-$PemK_{Sa}$ was lowest among all tested constructs. The influence of the linker type on synthesis level was different for the constructs containing GFP at the C-terminus and those containing GFP at the N-terminus. In the former group, the highest production was obtained using the α-helical linker while in the latter of the flexible one. Concerning solubility, no pronounced differences

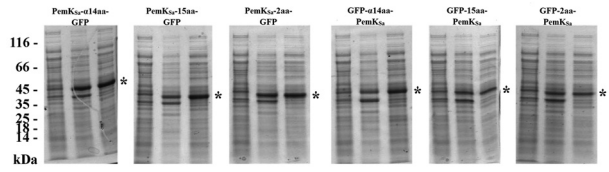

Figure 2. Production of evaluated fusion proteins (*) in *E. coli* BL21(DE3) as monitored by SDS-PAGE.
First lane in each panel represents the total cell lysate before IPTG induction; second and third lanes represent insoluble and soluble fraction after IPTG induction, respectively.

were observed between C-terminal fusions and those having GFP attached at the N-terminus. This does not corresponds to the previously published data suggesting that N-terminal fusion tags exhibit better synthesis and solubility (di Guan *et al.*, 1988). We observed no effect of the linker type on solubility of C-terminal GFP fusions, whereas among the N-terminal fusion constructs α-helical linker resulted in increased production of the fusion protein in the soluble fraction.

Assessment of fusion protein stability

To compare the stability and other properties of constructed fusion proteins each was purified using immobilized metal-affinity chromatography (IMAC) and gel filtration. First, we determined how the relative sequence of GFP and $PemK_{Sa}$ affects the efficiency of purification. For all constructs, an equal volume of the NiNTA was overloaded with the fusion protein. Overloading was evidenced by the presence of the fusion protein in the flow-through. After thorough washing, the proteins were eluted and the amount of fusion protein was quantified. In all tested cases higher amount of fusion protein was obtained for constructs containing GFP at the C-terminus (Fig. 3 and Table 3), indicating more efficient binding to the resin compared to constructs containing GFP at the N-terminus. We speculate that this effect may be related to steric constraints. In the configuration where the tag is linked to a smaller partner ($PemK_{Sa}$; ~12 kDa) more protein can be packed close to the resin than in the case when a larger partner (GFP; ~27 kDa) is direct-

Table 2. PCR primers used in the study.

Name	Sequence of the primer (5′-3′)*	Restriction site
GFP_BamHI-ForB	CTT**GGATCC**GAGTAAAGGAGAAGAACTTTTC	BamHI
GFP_PstI-RevB	AGT**CTGCAG**TTTGTATAGTTCATCCATG	PstI
PemK_PstI-ForB	TAT**CTGCAG**AACATTAAACAATTTGACATA	PstI
PemK_HindIII-RevB	AGT**AAGCTT**TTATAACGTTGGCTTAAGATA	HindIII
PemK_BamHI-For	TAT**GGATCC**GAACATTAAACAATTTGACATA	BamHI
PemK_PstI-Rev2	AGT**CTGCAG**TAACGTTGGCTTAAGATA	PstI
GFP_PstI-For	CTT**CTGCAG**AGTAAAGGAGAAGAACTTTTC	PstI
GFP_HindIII-Rev2	AGT**AAGCTT**TTATTTGTATAGTTCATCCATG	HindIII
GFP_PstI_LinH-RevB	AGT**CTGCAG**CGCTTTCGCGGCCGCTTCTTTCGCGGCCGCTTCGGCTTTGTATAGTTCATCCATG	PstI
GFP_PstI-LinH-For	CTT**CTGCAG**GCCGAAGCGGCCGCGAAAGAAGCGGCCGCGAAAGCGAGTAAAGGAGAAGAACTTTTC	PstI
GFP-PstI_LinE-RevB	AGT**CTGCAG**CGCGCTGCCACCGCCACCGCTGCCACCGCCACCGCTGGTTTTGTATAGTTCATCCATG	PstI
GFP_PstI_LinE-For	CTT**CTGCAG**ACCAGCGGTGGCGGTGGCAGCGGTGGCGGTGGCAGCGCGAGTAAAGGAGAAGAACTTTTC	PstI

*The sequence encoding the linker is underlined. The restriction sites are highlighted "bold"

Table 3. Quantitative evaluation of the proportion of fusion proteins in insoluble and soluble fractions and recovery during IMAC and dialysis.

Fusion protein	Total production (% of the highest value)	Fusion protein distribution [%]		Amount after IMAC using 1 ml of NiNTA resin (mg)	Recovery after dialysis (%)
		Insoluble fraction	Soluble fraction		
$PemK_{Sa}$-α14aa-GFP	100	49	51	32.5	67
$PemK_{Sa}$-15aa-GFP	79	40	60	30.3	79
$PemK_{Sa}$-2aa-GFP	79	44	56	44.5	8
GFP-α14aa-$PemK_{Sa}$	78	39	61	11.2	68
GFP-15aa-$PemK_{Sa}$	86	50	50	20.7	59
GFP-2aa-$PemK_{Sa}$	59	49	51	21.2	48

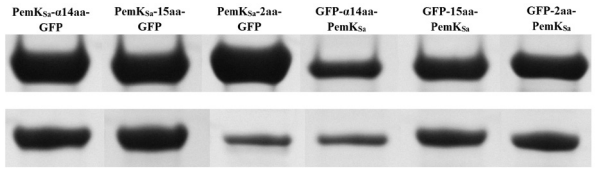

Figure 3. Semiquantitative SDS-PAGE analysis of samples before (upper panel) and after (lower panel) dialysis against low ionic strength buffer.

ly linked to the tag. This was however not further tested systematically.

Samples obtained after IMAC were dialyzed against low ionic strength buffer - one most suitable for further crystallization. It is important to note that our previous experience with His-tagged $PemK_{Sa}$ production demonstrated its instability in low ionic strength buffers, resulting in almost complete protein precipitation. Largely different recovery was observed between the analyzed constructs (Fig. 3 and Table 3). $PemK_{Sa}$-15aa-GFP was most resistant to low ionic strength and only 20% of the initial sample was lost during dialysis due to precipitation. $PemK_{Sa}$-2aa-GFP was most unstable in these conditions and over 90% of the initial sample was lost due to precipitation. The tested constructs containing GFP at the N-terminus of the fusion partner were relatively stable regardless the type of the linker. This demonstrates that the stability of the fusion protein in tested conditions is determined by a synergic effect of the relative position of the fusion partners and linker design.

Following dialysis, the multimeric state of obtained proteins was evaluated by gel filtration. Since gel filtration is routinely used as a polishing step in protein preparation we also investigated the recovery of particular constructs. Equal amounts of each fusion protein were separated on Superdex 75 equilibrated with low ionic strength buffer identical to that used for dialysis. The elution profiles differed significantly for N- and C-terminal fusions (Fig. 4). All fusion proteins containing GFP at their C-terminus eluted as a single peak corresponding to a fusion protein dimer. Most probably dimerization occurred through the GFP component, since $PemK_{Sa}$ was previously demonstrated in monomeric form in similar conditions (Bukowski *et al.*, 2013). Indeed, the crystal structure of GFP clearly demonstrates dimerization (Yang *et al.*, 1996). All the samples containing GFP linked at the N-terminus of $PemK_{Sa}$ presented a more heterogeneous gel filtration profile containing more than one protein peak. The two major peaks correspond to protein dimer and a monomer, but smaller peaks are additionally present which also contain the analyzed fusion protein, but the physicochemical character of which remains unknown. Clearly, dimerization is hindered in the N-terminal fusions, most probably due to a steric obstruction by the fusion partner being linked to the C-terminus of GFP, which part mediates dimerization (Yang *et al.*, 1996). This explanation is corroborated

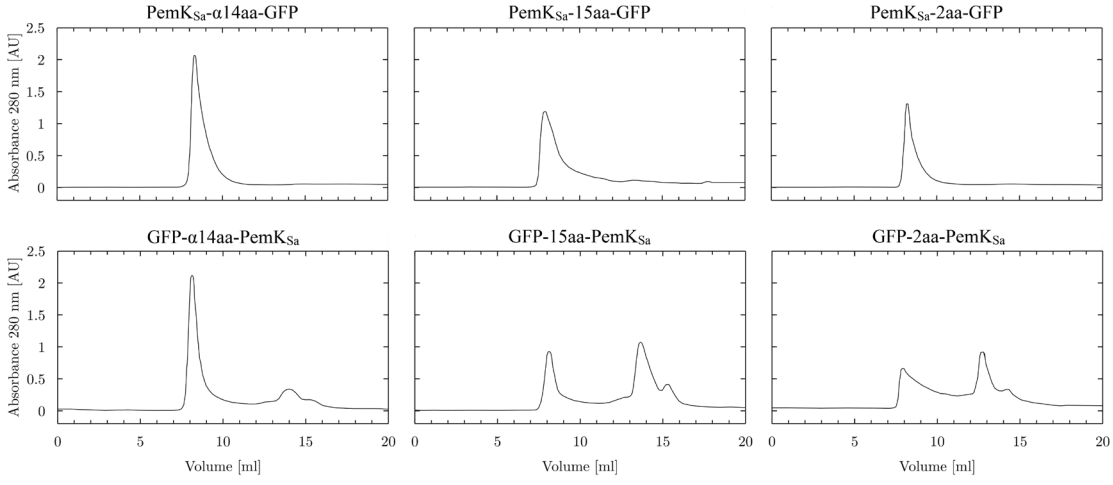

Figure 4. Gel filtration chromatograms of fusion proteins evaluated in this study.

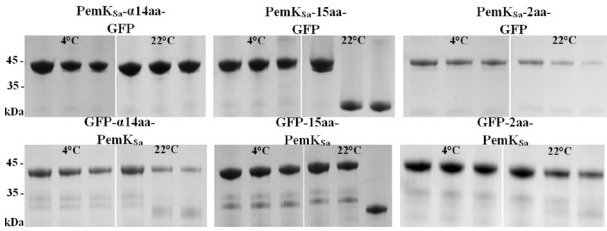

Figure 5. The effect of prolonged incubation on the stability of tested fusion proteins.
The fusion proteins were incubated in low ionic strength buffer for two weeks at 4°C and 22°C. The first lane in each panel represents the sample at time zero. The second and third lanes represent the samples after seven and fourteen day incubation, respectively.

by the fact that the fusion protein containing the long flexible linker presents much higher fraction of the dimer compared to the two other evaluated fusions. Interestingly, separation of GFP and PemK$_{Sa}$ with the flexible linker resulted in best recovery regardless the relative order of the evaluated fusion partners.

Protein crystallization usually involves prolonged incubation either at room temperature or at 4°C. To test the stability of obtained fusion proteins each was incubated in a buffer containing 5 mM Tris/HCl and 50 mM NaCl for two weeks. All tested variants of the fusion protein were stable for at least two weeks at 4°C. At room temperature (~22°C), however, significant degradation of all fusion proteins, save PemK$_{Sa}$-α14aa-GFP, was observed. Linker dependent differences in degradation pattern were noted. Proteins containing the short or the α-helical linker degraded slowly and without any stable intermediates indicating multiple sites of hydrolysis. In contrast, the proteins containing the flexible linker degraded much faster, but hydrolysis was observed only within the linker and PemK$_{Sa}$ while GFP remained stable during the time of the experiment (Fig. 5).

Concluding, our study demonstrates the influence of relative position of GFP and the fusion partner as well as of the linker length and design on production yield, soluble fraction content, efficiency of affinity purification and stability both during dialysis as well as during prolonged incubation. Among the examined proteins, PemK$_{Sa}$-α14aa-GFP exhibited all the properties desirable for a crystallization candidate. It was efficiently produced, well behaved during purification and stable over time. Of more general conclusions, we showed that C-terminal fusions are advantageous during gel filtration due to homogenous elution profile. We have moreover shown that short and helical linkers are more resistant to hydrolysis. These findings are important for GFP containing fusion protein design.

Acknowledgements

Faculty of Biochemistry, Biophysics and Biotechnology is a partner of the Leading National Research Center (KNOW) supported by the Ministry of Science and Higher Education.

Acknowledgements of financial support

This work was supported in part from the grant UMO-2014/13/B/NZ1/00043 (to BW) from the National Science Centre, Poland. We acknowledge the financial support of European Union structural funds (grant POIG.02.01.00-12-167/08).

Conflicts of Interest

We declare no conflict of interest.

REFERENCES

Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng* **14**: 529–532.

Ashikawa Y, Ihara M, Matsuura N, Fukunaga Y, Kusakabe Y, Yamashita A (2011) GFP-based evaluation system of recombinant expression through the secretory pathway in insect cells and its application to the extracellular domains of class C GPCRs. *Protein Sci* **20**: 1720–1734.

Bukowski M, Lyzen R, Helbin WM, Bonar E, Szalewska-Palasz A, Wegrzyn G, Dubin G, Dubin A, Wladyka B (2013) A regulatory role for *Staphylococcus aureus* toxin-antitoxin system PemIK$_{Sa}$. *Nat Commun* **4**: 2012.

Charpentier E, Anton AI, Barry P, Alfonso B, Fang Y, Novick RP (2004) Novel cassette-based shuttle vector system for gram-positive bacteria. *Appl Environ Microbiol* **70**: 6076–6085.

Cherezov V, Rosenbaum DM, Hanson MA, Soren GF, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science* **318**: 1258–1265.

Corsini L, Hothorn M, Scheffzek K, Sattler M, Stier G (2008) Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein Sci* **17**: 2070–2079.

Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY (1995) Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci* **20**: 448–455.

Di Guan C, Li P, Riggs PD, Inouye H (1988) Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion maltose-binding protein. *Gene* **67**: 21–30.

Hsieh JM, Besserer GM, Madej MG, Bui HQ, Kwon S, Abramson J (2010) Bridging the gap: A GFP-based strategy for overexpression and purification of membrane proteins with intra and extracellular C-termini. *Protein Sci* **19**: 868–880.

Iwata S, Ostermeier C, Ludwig B, Michel H (1995) Structure at 2,8 A resolution of cytochrome c oxidase from *Paracoccus denitrificans*. *Nature* **376**: 660–669.

Japrung D, Chusacultanachai S, Yuvaniyama J, Wilairat P, Yuthavong Y (2005) A simple dual selection for functionally active mutants of *Plasmodium falciparum* dihydrofolate reductase with improved solubility. *Protein Eng Des Sel* **18**: 457–464.

Kobe B, Center RJ, Kemp BE, Poumbourios P (1999) Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. *Proc Natl Acad Sci USA* **96**: 4319–4324.

Kwolik CM, Hengstenberg W (2001)Use of *Staphylococcus aureus* 6-P-galactosidase and GFP as fusion partners for lactose-specific IIC domain from *Staphylococcus aureus*. *J Mol Microbiol Biotechnol* **3**: 395–400.

Lally JM, Newman RH, Knowles PP, Islam S, Coffer AI, Parker M, Freemont PS (1998) Crystallization of an intact GST-estrogen receptor hormone binding domain fusion protein. *Acta Cryst* **D54**: 423–426.

Mueller GA, Pedersen LC, Lih FB, Glesner J, Moon AF, Chapman MD, Tomer KB, London RE, Pomés A (2013) The novel structure of the cockroach allergen Bla g 1 has implications for allergenicity and exposure assessment. *J Allergy Clin Immunol* **132**: 1420–1426.

Nguyen HB, Hung LW, Yeates TO, Terwilliger TC, Waldo GS (2013) Split green fluorescent protein as a modular binding partner for protein crystallization. *Acta Cryst D Biol Crystallogr* **69**: 2513–2523.

Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov* **5**: 993–996.

Robinson CR, Sauer RT (1998) Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proc Natl Acad Sci U S A* **95**: 5929–5934.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**: 671–675.

Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein Sci* **12**: 1313–1322.

Suzuki N, Hiraki M, Yamada Y, Matsugaki N, Igarashi N, Kato R, Dikic I, Drew D, Iwata S, Wakatsuki S, Kawasaki M (2010) Crystallization of small proteins assisted by green fluorescent protein. *Acta Cryst D Biol Crystallogr* **66**: 1059–1066.

Takeuchi S, Kinoshita T, Kaidoh T, Hashizume N (1999) Purification and characterization of protease produced by *Staphylococcus aureus* isolated from a diseased chicken. *Vet Microbiol* **67**: 195–202.

Yang F, Moss LG, Philips GN (1996) The molecular structure of green fluorescent protein. *Nat Biotechnol* **14**: 1246–1251.