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 PemKαα 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. PemKαα-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 partner 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 vitro 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 **A preliminary report on the same subject was presented at the XLI 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.
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\textsubscript{Sa} (R84A). The interferase, belonging to the toxin-antitoxin system encoded on pCH91 plasmid of \textit{S. aureus} CH91 (Takeuchi et al., 2012), 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.** \textit{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. \textit{pemK\textsubscript{Sa}} (following denoted as \textit{pemK\textsubscript{Sa}}) was amplified from pETDuet-\textit{pemK\textsubscript{Sa}} (Bukowski et al., 2013). \textit{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 Tryp-tic Soy Broth (Sigma Aldrich) containing ampicillin (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 21000×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 PemK\textsubscript{Sa} of \textit{S. aureus} strain CH91 (Bukowski et al., 2013), however, our attempts to obtain recombinant PemK\textsubscript{Sa} or its mutants at crystallization quality have
A systematic investigation of the stability of GFP fusion proteins

The sequence encoding the linker is underlined. The restriction sites are highlighted “bold.”

**Table 2. PCR primers used in the study.**

| Name            | Sequence of the primer (5'-3')* | Restriction site |
|-----------------|---------------------------------|------------------|
| GFP_BamHI-ForB  | AGTGGATCCCTTCAGTATTCATCCATG     | BamHI            |
| GFP_BamHI-RevB  | CTGGGATCCCAAGAGAGGAGAGAATTTCC   | BamHI            |
| PemK_PstI-ForB  | TATCTGAGAACATTAAACAACATTTGACATA| PstI             |
| PemK_PstI-RevB  | AGTGAAGCTTATTTAAGGTGCCATTTGACTA | PstI             |
| PemK_HindIII-For| TATGGATCCGAACATTAAACAACATTTGACATA | HindIII          |
| PemK_HindIII-RevB | AGTGAAGCTTATTTAAGGTGCCATTTGACTA   | HindIII          |
| GFP_PstI-For    | AGTTCAAGCTTATTTAAGGTGCCATTTGACTA | PstI             |
| GFP_PstI-Rev    | AGTGGATCCCTTCAGTATTCATCCATG     | PstI             |
| GFP_HindIII-For | TATGGATCCGAACATTAAACAACATTTGACATA | HindIII          |
| GFP_HindIII-RevB | AGTGAAGCTTATTTAAGGTGCCATTTGACTA   | HindIII          |
| GFP_PstI_LinH-For | AGTTCAAGCTTATTTAAGGTGCCATTTGACTA | PstI             |
| GFP_PstI_LinH-RevB | AGTGGATCCCTTCAGTATTCATCCATG     | PstI             |
| GFP_PstI_LinE-For | AGTTCAAGCTTATTTAAGGTGCCATTTGACTA   | PstI             |
| GFP_PstI_LinE-RevB | AGTGGATCCCTTCAGTATTCATCCATG     | PstI             |

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

Figure 1. Schematic representation of fusion proteins evaluated in this study.

PemKα 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 PemKα (R84A) (further denoted as PemKα) 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 PemKα 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 PemKα-2aa-GFP whereas production level of GFP-2aa-PemKα 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 observed using the α-helical linker while in the latter of the flexible former group, the highest production was obtained using those containing GFP at the N-terminus of PemKα and Table 3). Highest production 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 (PemKα, ~12 kDa) more protein can be packed close to the resin than in the case when a larger partner (GFP, ~27 kDa) is direct.

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 PemKα 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 (PemKα, ~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.**
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 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 by the data presented in Table 3.

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 NITA resin (mg) | Recovery after dialysis (%) |
|---------------------------------|-------------------------------------------|---------------------------------|-----------------------------------------------|-----------------------------|
| PemK$_{sa}$-α14aa-GFP       | 100                                       | Insoluble fraction: 49          | Soluble fraction: 51                          | 32.5                         | 67                           |
| PemK$_{sa}$-15aa-GFP       | 79                                        |                                 |                                               |                             | 30.3                         | 79                           |
| PemK$_{sa}$-2aa-GFP       | 79                                        |                                 |                                               | 44.5                         | 8                            |
| GFP-α14aa-PemK$_{sa}$       | 78                                        |                                 |                                               | 11.2                         | 68                           |
| GFP-15aa-PemK$_{sa}$       | 86                                        |                                 |                                               | 20.7                         | 59                           |
| GFP-2aa-PemK$_{sa}$       | 59                                        |                                 |                                               | 21.2                         | 48                           |

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

Figure 4. Gel filtration chromatograms of fusion proteins evaluated in this study.
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 di- mer compared to the two other evaluated fusions. Interest- ingly, separation of GFP and PemKα with the flexible linker resulted in best recovery regardless the relative or- der of the evaluated fusion partners.

Protein crystallization usually involves prolonged incuba- tion 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 PemKα-z14aa-GFP, was observed. Linker dependent differences in degradation pattern were noted. Proteins containing the short or the z-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 PemKα 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, PemKα-z14aa-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 contain- ing fusion protein design.

Acknowledgements

Faculty of Biochemistry, Biophysics and Biotechnol- ogy is a partner of the Leading National Research Cent- er, 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, Kaniy N, Nagamune T (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng 14: 520–532.

Ashikawa Y, Ibara M, Matsura 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, Szałowski-Palasz A, Wegrzyn G, Dubin G, Dubin A, Władyka B (2013) A regulatory role for Staphylococcus aureus toxin-antitoxin system PemRS. Nat Commun 4: 2012.

Charpentier E, Anton AI, Barry P, Alfonso B, Fang Y, Viovyck RP (2004) Novel cassette-based shuttle vector system for gram-positive bacteria. Appl Envir Microb 70: 6676–6685.

Cherezo V, Rosenbaum DM, Hanson MA, Soren GF, Rasmussen SG, Thian FS, Kohlbilla TS, Choi HJ, Kuhn P, Wei W, Kohlbilla B, Stevens RC (2007) High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. Science 318: 1258–1265.

Corisini L, Hothorn M, Scheffzick K, Sattler M, Stier G (2008) Thiore- doxin as a fusion tag for carrier-driven crystallization. Protein Sci 17: 2070–2079.

Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tisen RV (1995) Understanding, improving and using green fluorescent proteins. Trends Biochem Sci 20: 445–455.

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

Hsieh JM, Besserer GM, Malej MG, Hui 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: 686–880.

Iwata S, Ostermeier C, Ludwig B, Michel H (1995) Structure at 2.8 Å resolution of cytochrome c oxidase from Paracoccus denitrificans. Nature 376: 668–669.

Japung D, Chusakutanacai S, Yuvaniyama J, Willairat P, Yuthavong Y (2005) A simple dual selection for functionally active mutants of Plasmodium falciparum dhfr/dhfrts reduceada with improved solubil- ity. Protein Eng Des Sel 18: 457–464.

Kobe B, Center RJ, Kemp BE, Ponzoubaris P (1999) Crystal struc- ture of human T cell leukemia virus type 1 gp110 ectodomain is stabilized as a maltose-binding protein chimera reveals structural evo- lution of retroviral transmembrane proteins. Proc Natl Acad Sci U S A 96: 4319–4324.

Kobayashi YM, Hengstenberg W (2001) Use of Staphylococcus aureus 6-P-galactosidase and GFP as fusion partners for lactose-specific IIC doxir: A 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 D59: 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, Sau R (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, Rashband WS, Eliee-CI EW (2012) NIH Image to Im- age: 25 years of image analysis. Nat Rev Mol Cell Biol 9: 415–426.