Selecting an appropriate method for expressing S locus F-box-S2 recombinant protein

Jahanshah Ashkania,b,⁎, D.J.G. Reesa,b

a Biotechnology Department, University of the Western Cape, Robert Sobukwe Road, Bellville, 7535, South Africa
b Agricultural Research Council, Biotechnology Platform, Private Bag X5, Onderstepoort, 0110, South Africa

ARTICLE INFO

Keywords:
S locus F-box-S2
Recombinant protein
Protein expression

ABSTRACT

A single locus (S locus) including at least two linked genes (female and male determinants) genetically controls the gametophytic self-incompatibility (GSI) in apple, which has evolved to avoid self-fertilization. There has been extensive work done on the female determinant of self-incompatibility, which has led to the determination of the tertiary structure of S-RNase. However, the tertiary structure of male determinant (S locus F-box, SLF/SFB) remains unresolved, which could mainly be due to difficulties associated with its expression in the recombinant expression systems. In addressing this, we have evaluated several in vivo (prokaryotic and eukaryotic) and in vitro expression systems for their efficiency in the expression of apple SLF2. The most successful expression of SLF2 (1 mg/ml) was achieved in E. coli using the synthesized gene in a high salt culture and applying heat shock before induction of culture. We therefore present an approach for the efficient expression of S locus F-box recombinant proteins for future functional and structural studies.

1. Introduction

The growing economic importance of fruit crops and the problems associated with their commercial production has placed the molecular basis of self-incompatibility (SI) under much attention. There has been extensive work done on the female determinant of self-incompatibility, which has led to the determination of the tertiary structure of S-RNase in Pyrus pyrifolia [1,2]. However, the male determinant (S locus F-box, SLF/SFB) structure remains unresolved, which could mostly be due to difficulties related to its expression in the recombinant expression systems.

Although, gene expression is a strong tool for studying biological processes, choosing an appropriate expression system based on the characteristics and deliberated application of a recombinant protein is critical to obtain adequate quantities of the protein in a timely manner [3,4]. Many expression systems are available but no single system has yet been developed that works for all proteins [5–11].

Escherichia coli is a popular organism in in vivo studies due to its relatively easy and inexpensive genetic manipulation, scalability of the experiments, low cost, fast growth, high-density cultivation and its ability to express labeled proteins [5,9,12]. However, certain limitations may be experienced primarily at the initiation of translation [13–17]. To overcome such limitations certain E. coli strains can over-express rare tRNAs (for AGG, AGA, CUU, AUA, CCC, and GGA codons) and hence significantly improve gene expression [18–20]. Furthermore, genes are designed in certain cases to match host bias or to maximize Codon Adaptation Index (CAI) and contribute towards an improved expression [13,21]. In addition, synthetic gene sequences designed by codon engineering can create an optimal heterologous gene expression and are considered important in strategically improve protein expression [22,23].

Besides E. coli, several other recombinant protein expression systems are available in eukaryotes, including fungal, mammalian, amphibian, insect, plants and yeast cells. Among the available host expression systems, Escherichia coli and cultivated mammalian cells are considered most widely used [24–26]. Advantages of using a mammalian expression system include its ability for post-translational modifications (PTMs) such as correct protein folding and glycosylation. The protein yield of mammalian expression systems can be optimized by the use of right cell line, expression vector, promoter elements and transfection efficiency of the mammalian cells [25–27].

Another commonly used and well-established technique in recombinant protein expression is the Transient Expression Arabidopsis Mesophyll Protoplasts (TEAMP), which is appropriate for studying various physiological processes. The TEAMP is a highly useful and sensitive system because it does not need a large number of protoplasts, sterile techniques or complex culture medium [28].

Additional to in vivo expression systems, cell-free (in vitro) protein
expression methods have become good alternatives for cell-based protein production. Firstly, it is easily adapted for high-throughput and automated procedures. Secondly, the availability of various prokaryotic and eukaryotic cell lysates provides a flexible choice for high-level protein expression, especially those that are difficult to express under an in vivo system. Finally, cell-free systems are capable of synthesizing large protein–protein complex mixtures in a single reaction, which makes them an exploitable system for developing proteomic tools [29–32]. Cell-free protein synthesis has many applications in the fields of genomics, proteomics, and synthetic biology while allowing for increased productivity and various post-translational modifications [33,34],

Despite great progress in understanding the complex non-self recognition between male and female determinants, a comprehensive structural study is still necessary to understand the biochemical basis of differential interactions between male and female determinants (i.e. SLF/SFB and SRNase) in Rosaceae. Here, we aim to evaluate the capability and applicability of several prokaryotic and eukaryotic expression systems in expressing the S locus F-box-S2 recombinant protein. Since F-box proteins represent a big family of proteins in plants (e.g., the Arabidopsis thaliana genome encodes more than 700 F-box proteins) and multiple F-box genes with more than 60% similarity have been recognized at the S-loci of the families that own S-RNase-based SI [35–37], the information provided here not only improves our understanding of SRNase-based self-incompatibility, but also lay the foundation for the functional and structural studying of the interactions between F-box proteins and their substrates.

2. Material and methods

2.1. Recombinant protein expression in E. coli and purification

For the purpose of construct preparation for recombinant protein expression; nucleic acid extraction, PCR amplification, and DNA sequencing were done (Supplementary information 1). The SLF2-pGEX-6P-2, SLF2-pET21a and SLF2-pETM-41 constructs were prepared by digestion of SLF2-pGEM-T Easy constructs using double digestion with BamHI and XhoI, and NcoI and BamHI respectively, where the restriction enzyme sites were added to SLF2 codon sequence (Accession Number: DQ4422811) using designed forward and reverse primers (Table S1) by PCR technique. The restriction enzyme assays were used to check the availability of SLF2 in the constructs. Separate conditions were considered with combining vector plasmids, cell lines and IPTG concentrations (0.1 mM–1.0 mM), while the temperature was decreased to (25–30 °C) following IPTG induction and cells were harvested between 4 and 24 h with 4 h intervals. The absence of SLF2 expression was further assessed using blotting techniques (Supplementary information 2.1). Gene synthesis was carried out by GenScript (http://www.genscript.com) after a wide variety of factor optimization using OptimumGeneTM algorithm to generate a single sequence gene that can reach the highest possible level of expression in E. coli (Supplementary information 2.2, Fig. S1 and S2). The localization of the GST-SLF2opt in E. coli was analyzed by using osmotic shock procedure and an anionic detergent (Sarkosyl) and 8 M Urea. GST-SLF2opt was expressed in E. coli C41 (DE3) pLysS while grown in LB containing 0.5 M NaCl and induced with 0.1 mM IPTG. Furthermore, solubility and localization of GST-SLF2opt were assessed (Supplementary information 2.3).

2.2. Recombinant protein expression in mammalian cells

The SLF2-FX-6xHis-pCAMBI1-305.1 was constructed by the digestion of SLF2-pGEM-T-Easy constructs using double digestion of Ndel and BamHI restriction enzymes whose sites were added to SLF2 (Table S1) using PCR technique. SLF2 expression was conducted using pCAMBI1305.1 expression vector as a 6xHis fusion protein with a designed Factor Xa cleavage site before the 6xHis tag at the C-terminal end (Table S1). The PEG-mediated method was used to transfect a large amount of SLF2-FX-6xHis-pCAMBI1305.1 plasmid DNA (30–40 μg) into the healthy and successful isolated protoplasts. Subsequently, the samples were loaded on the SDS-PAGE and the Western blot technique was applied to confirm the expression of SLF2 in the protoplasts (Supplementary information 4).

2.3. Recombinant protein expression in Arabidopsis thaliana protoplasts

The SLF2-FX-6xHis-pCAMBI1-305.1 was constructed by the digestion of SLF2-pGEM-T-Easy constructs using double digestion of Ndel and BamHI restriction enzymes whose sites were added to SLF2 (Table S1) using PCR technique. SLF2 expression was conducted using pCAMBI1305.1 expression vector as a 6xHis fusion protein with a designed Factor Xa cleavage site before the 6xHis tag at the C-terminal end (Table S1). The PEG-mediated method was used to transfect a large amount of SLF2-FX-6xHis-pCAMBI1305.1 plasmid DNA (30–40 μg) into the healthy and successful isolated protoplasts. Subsequently, the samples were loaded on the SDS-PAGE and the Western blot technique was applied to confirm the expression of SLF2 in the protoplasts (Supplementary information 4).

2.4. Cell-free (in vitro) recombinant protein expression

The pIVEX (In Vitro EXpression) vectors are specially designed for gene expression in vitro. These vectors contain an ampicillin resistance gene, T7 promoter, and 6xHis-tag that facilitate the high level of expression, detection and purification of proteins. The pIVEX2.3d (3560 bp) contains an un-cleavable 6xHis-tag at the C-terminal end. The pIVEX2.4d (3583 bp) contains a 6xHis-tag at the N-terminal end, which is cleavable with Factor Xa. The SLF2-pIVEX2.3d and SLF2-pIVEX2.4d were constructed by the cloning of SLF2 gene isolated from apple var. Golden Delicious into pIVEX2.3d using NcoI and SmaI and pIVEX2.4d expression vector using NotI and SmaI restriction sites added to the SLF2 codon sequence (Table S1) using PCR technique, respectively. To find the best condition for maximum expression per reaction both constructs of SLF2-pIVEX2.3d and SLF2-pIVEX2.4d were used as templates in concentrations of 5 μg/ml, 8.5 μg/ml and 10 μg/ml along with 0.27% (v/v) of cell extract and standard reaction mix. Samples were incubated at temperatures of 25, 30 and 37 °C in an Eppendorf Thermomixer R for 4–12 h while shaking at 250 rpm. For the protein expression assessment, samples were centrifuged at 10,000 × g for 15 min and 50 μl of suspension was loaded on the SDS-PAGE and analyzed by Western blotting (Supplementary information 5). When the incubation time increased from 4 to 12 h, the expression of the SLF2 gene was increased accordingly (data not shown).

Results

3.1. Recombinant protein expression in E. coli

The main focus of this study has been the over-expression of S locus F-box-S2 pollen factor of apple var. Golden Delicious (SLF2) in a bacterial expression system to provide sufficient amount of SLF2 for protein studies. To achieve this goal, several cloning vectors (pGEX-6P-2, pET21a and pETM-41) were tested in different bacterial cell lines (BL21(DE3)-pLysS, C41(DE3)-pLysS, and C43(DE3)-pLysS). However, no significant expression for SLF2 was found.
3.2. Recombinant protein expression in mammalian cells

SLF2 expression was conducted using pcDNA3.1/Zeo(+) expression vector as a 6xHis fusion protein with a designed Factor Xa cleavage site before the 6xHis tag at the C-terminal end (Table S1). CHO cells were grown in complete Ham's F12 at 37 °C in a humidified atmosphere containing 5% CO₂. The results of this study proved successful although only very faint expression of this gene was observed in the CHO cells lysates (Fig. 1).

3.3. Recombinant protein expression in mesophyll protoplasts of Arabidopsis

We further examined Arabidopsis protoplasts as an alternative expression host for expressing the plant-derived SLF2 protein. Despite the successful use of protoplast expression system by other scientists, the results indicate a very faint signal on the PVDF membrane at 45 kDa (Fig. 2) while no detectable band was observed on the SDS-PAGE suggesting a very low level of expression of the SLF2 protein.

3.4. Cell free (in vitro) recombinant protein expression

To find the best condition for maximum expression per reaction both constructs of SLF2-pIVEX2.3d and SLF2-pIVEX2.4d were used as templates in several DNA concentrations (5–10 μg/ml) under temperatures between 25 and 37 °C for 4–12 h. The result of the optimization shows that the concentration of 8.5 μg/ml DNA in reaction while incubated at 30 °C for 12 h, yields the best results for the expression of SLF2 in vitro (Fig. 3). Nonetheless, the in vitro expression system did not meet our priorities due to its shortcomings such as it being costly and its low protein yield (~ 25–30 μg/ml).

3.5. Recombinant protein expression using synthetic SLF2 gene (SLF2opt) in E. coli

We carried out SLF2 gene synthesis through GenScript (http://www.genscript.com) after a broad range of factors were optimized using OptimumGeneTM algorithm to generate a single sequence gene that can reach the highest possible level of expression in E. coli. The SLF2opt-pGEX-6P-2 was constructed using BamHI and XhoI restriction enzyme sites and the construct was transformed into the C41 (DE)-pLysS cells. The growth of bacterial cultures containing SLF2opt-pGEX-6P-2 was measured at OD600 every 30 min after the IPTG induction and the result showed that the GST-SLF2opt protein inhibits the growth of E. coli cells (data not shown) and an over-expression of GST-SLF2opt fusion protein (~ 70 kDa) was shown in the insoluble fraction (Fig. 4).

To enhance the solubility of GST-SLF2opt fusion protein, growth medium (LB) was replaced with a high salt growth medium (LB containing 0.5 M NaCl) and a heat-shock step (incubating the culture in a 47 °C water-bath for 30 min) was applied before IPTG induction when the OD600 of the culture was between 0.8-1. Sixteen different parameters containing IPTG concentrations (0.1, 0.25, 0.5 and 1 mM) and overnight (16 h) incubation temperatures (18, 20, 25 and 30 °C) were examined using transformed BL21, BL21-CodonPlus, C41 (DE3)-pLysS and C43 (DE3)-pLysS cell lines. The Dot blot immunoassay was used as a technique for rapid testing of the parameters for maximum expression of GST-SLF2opt fusion. The result of the Dot blot assay shows a significant signal for C41 (DE3)-pLysS and C43 (DE3)-pLysS cell lines when compared to BL21 and BL21-CodonPlus. Furthermore, the most significant signal belongs to C41 (DE3)-pLysS cells that were induced with 0.1 mM IPTG concentration and further incubated at 20 °C overnight (Fig. 5).
3.5.1. Cellular localization of GST-SLF2opt fusion protein

The localization of the GST-SLF2opt in E. coli was assessed using osmotic shock, anionic detergent (Sarkosyl) and 8 M Urea. GST-SLF2opt was expressed in E. coli C41 (DE3)-pLysS while grown in LB containing 0.5 M NaCl and induced with 0.1 mM IPTG. This resulted in the yield of 1 mg/l of soluble protein (proteins in periplasmic, inner membrane and cytoplasmic fractions) measured at A260. However, the biggest amount of expressed proteins was presented in the form of inclusion bodies (Fig. 6). We further used a GST column for affinity purification of the GST-SLF2opt recombinant protein. To release the GST domain from GST-SLF2opt, PreScission protease was used in a dialysis bag while stirring in 1 litter cleavage buffer at 4 °C overnight. Following the cleavage, about three-quarter of SLF2opt proteins was precipitated in the absence of the fusion tag in the dialysis bag (data not shown) and the amount of soluble SLF2opt was decreased to 0.25 mg/l.

Fig. 7 illustrates the elution fraction containing GST-SLF2opt fusion protein (~ 70 kDa) obtained from the protein purification step before and after 3C protease cleavage. Two bands in Lane C correspond to GST (26.6 kDa) and the SLF2opt (43 kDa) proteins, which is indicative of the complete cleavage of GST-SLF2opt fusion protein.

We used Anion exchange chromatography with a pre-packed MonoQ 10/100 GL column on AKTA Chromatography System (GE Healthcare Life Sciences) to separate GST proteins from SLF2opt recombinant protein in the cleaved sample. The chromatogram shows a number of small peaks containing SLF2opt recombinant protein and a big peak containing GST proteins as confirmed on an SDS-PAGE (Fig. 8).

Discussion

It is postulated that the lack or slow progress in structural aspects of the Rosaceae self-incompatibility may be due to the existing limitations on the expression of SLF/SFB in the heterologous expression systems. Accordingly, we evaluated several expression systems for their efficiency in expressing adequate quantity of the S locus F-box-S2 (SLF2) recombinant protein. Most of these methods had limitations regarding their yield, cost, and time efficiency. However, the successful expression of SLF2 was achieved in E. coli using the synthesized gene. Furthermore, SLF2 was expressed in a soluble form using high salt culture and heat-shock application before the induction of culture.

Here we report that the production of an SLF2 protein optimized for expression in an E. coli-based system is shown to be the most desirable technique for high yield of SLF2. Using a high-salt media along with heat shock regimes before induction significantly contributed to the production of more soluble recombinant proteins even though a large proportion of GST-SLF2opt fusion protein was still in the inclusion bodies. This necessitates future studies to focus on the optimization of the conditions that give the maximum amount of GST-SLF2opt in a soluble form. Additionally, more work needs to be carried out to optimize the purification of SLF2, as we encountered protein precipitation during the purification steps.

In a study on Antirrhinum (family: Plantaginaceae) Qiao et al. showed that the expression of AhSLF-S2 was only made possible following the separation of the protein into 6xHis-AhSLF-S2-N and 6xHis-AhSLF-S2-C for the pull-down assay [38]. However, it is well established that the tertiary structure of a protein is essential in understanding its function [39]. We, therefore, aimed at expressing SLF2 as a complete protein molecule, however despite the use of several E. coli expression systems, no significant expression was observed. This lack of expression in E. coli could be due to un-optimized factors responsible for the efficiency of heterologous protein production such as biased codon usage and mRNA secondary structure [12,40]. As a result, mammalian cell expression systems were investigated as transient expression systems that allow extra-chromosomal amplification of plasmids and hence permit more plasmid copies to persist in the transfected cells during the production phase for a significant increase in gene expression. The large-scale expression of recombinant proteins in mammalian cells has been shown to generate sufficient amounts of recombinant protein for early production and high-throughput screening in functional studies [41].
However, the expression of SLF/SFB recombinant proteins in mammalian cells requires optimization of the technique to improve the transfection procedure through increasing speed and reducing the costs, all of which can possibly make mammalian cells more suitable for SLF/SFB recombinant protein expression. We further assessed the use of protoplasts isolated from plants as means for the transient expression of recombinant proteins. Protoplasts retain their cell identity and show high transformation efficiency while requiring low maintenance [42].

Despite the successful use of protoplast expression system as a recombinant expression system by other scientists such as Wu et al., very faint expression of the SLF2 protein was observed in this study.

Another system examined in this study is the cell-free/in vitro expression system that can synthesize proteins in a high speed and accurate manner. The results obtain from this analysis suggest very low levels of expression. The in vitro system, by eliminating the need to manipulate whole cells in culture is a suitable technique for automated, high-throughput expression of a large number of proteins. However, the very low productivity i.e. less than 30 μg protein/ml reaction of this system limits its applicability and make it unfeasible compared to cell-based systems.

Subsequently, the optimization of the E. coli system to express SLF/SFB gene was considered using artificial gene synthesis. Considering that each organism carries its own bias in the usage of the 61 available codons [15] in some cases expression is limited primarily to the initiation of translation [13,14,43,44]. In addition to codon bias, mRNA secondary structures are identified which affect the expression of the encoded proteins [15]. A synthetic gene sequence by codon engineering is an important strategy for enabling optimal heterologous gene expression through making DNA templates ready to be expressed in any heterologous host [22,23]. This method verified to significantly contribute towards the expression of SLF/SFB in E. coli. As such it may be deduced that among the expression systems under study artificial gene synthesis optimized for over-expression in E. coli provides the best technique for the optimal expression of SLF/SFB and may further be used in future structural studies while folding and activity of the

(a) SDS-PAGE of the total bacterial lysates. (b) Western blot analysis of the total bacterial lysates.
expressed protein must be evaluated. Having said that, the protocols provided for the mammalian cells, Arabidopsis protoplast and expressed protein must be evaluated. Having said that, the protocols

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.btre.2017.06.005.
scale production of functional mammalian membrane proteins suitable for X-ray crystallography, J. Mol. Microbiol. Biotechnol. 18 (2) (2010) 85–91.

[35] J.M. Gagne, B.P. Downes, S.H. Shiu, A.M. Durski, R.D. Vierstra, The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in Arabidopsis, Proc. Natl. Acad. Sci. U. S. A. 99 (17) (2002) 11519–11524.

[36] E.P. Risseeuw, T.E. Daskalchuk, T.W. Banks, E. Liu, J. Cotelesage, H. Hellmann, M. Estelle, D.E. Somers, W.L. Crosby, Protein interaction analysis of SCF ubiquitin E3 ligase subunits from Arabidopsis, Plant J. 34 (6) (2003) 753–767.

[37] Z. Hua, X. Meng, T.-h. Kao, SLF-like proteins reveals its unique function in S-RNase?based self-incompatibility, Plant Cell 19 (11) (2007) 3593–3609.

[38] H. Qiao, F. Wang, L. Zhao, J. Zhou, Z. Lai, Y. Zhang, T.P. Robbins, Y. Xue, The F-box protein AhSLF-S2 controls the pollen function of S-RNase-based self-incompatibility, Plant Cell 16 (9) (2004) 2307–2322.

[39] M.F. Perutz, Species adaptation in a protein molecular biosystemsecule, Mol. Biol. Evol. 1 (1) (1983) 1–28.

[40] E.L. Rashdy, M. Redwan, The optimal gene sequence for optimal protein expression in Escherichia coli: principle requirements, Arab J. Biotechnol. 9 (3) (2006) 493–510.

[41] R. Kunaparaju, M. Liao, Sunstrom N-A: Epi-CHO, an episomal expression system for recombinant protein production in CHO cells, Biotechnol. Bioeng. 91 (6) (2005) 670–677.

[42] J. Sheen, Signal transduction in maize and Arabidopsis mesophyll protoplasts, Plant Physiol. 127 (4) (2001) 1466–1475.

[43] M. Welch, A. Villalobos, C. Gustafsson, J. Minshull, You’re one in a googol: optimizing genes for protein expression, J. R. Soc. Interface 6 (Suppl. 4) (2009) 467–476.

[44] M. Welch, S. Govindarajan, J.E. Ness, A. Villalobos, A. Gurney, J. Minshull, C. Gustafsson, Design parameters to control synthetic gene expression in Escherichia coli, PLoS One 4 (9) (2009) e7002.