Protein Expression and Purification xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Tandem SUMO fusion vectors for improving soluble protein expression and purification

Fernando Guerrero, Annika Ciragan, Hideo Iwaï *

Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, P.O. Box 65, Helsinki FIN-00014, Finland

ARTICLE INFO

Article history:
Received 22 June 2015
and in revised form 17 August 2015
Accepted 18 August 2015
Available online xxxx

Keywords:
Fusion proteins
Solubility enhancement tag
SUMO system
Ulp1 protease
TonB
Scytovirin
Protein ligation

ABSTRACT

Availability of highly purified proteins in quantity is crucial for detailed biochemical and structural investigations. Fusion tags are versatile tools to facilitate efficient protein purification and to improve soluble overexpression of proteins. Various purification and fusion tags have been widely used for overexpression in *Escherichia coli*. However, these tags might interfere with biological functions and/or structural investigations of the protein of interest. Therefore, an additional purification step to remove fusion tags by proteolytic digestion might be required. Here, we describe a set of new vectors in which yeast SUMO (SMT3) was used as the highly specific recognition sequence of ubiquitin-like protease 1, together with other commonly used solubility enhancing proteins, such as glutathione S-transferase, maltose binding protein, thioredoxin and trigger factor for optimizing soluble expression of protein of interest. This tandem SUMO (T-SUMO) fusion system was tested for soluble expression of the C-terminal domain of TonB from different organisms and for the antiviral protein scytovirin.

1. Introduction

Although the use of recombinant proteins has been a valuable advance in recent times, the choice of the appropriate host and expression system needs to be optimized on a case-by-case basis according to the target protein [1,2]. Purification tags like polyhistidine-tag are indispensable for facilitating efficient protein purification of heterogeneous proteins overexpressed in *Escherichia coli* [3,4]. In addition, various proteins have been used as fusion tags in combination with purification tags for improving properties like solubility and expression levels of target proteins [5]. Glutathione S-transferase (GST) [6], thioredoxin (TRX) [7], DsbA [8], maltose binding protein (MBP) [9], trigger factor (TF) [10] and others have been often used [11]. Various vectors for fusion proteins are already commercially available. However, fusion tags might interfere with biological assays or structural investigations, making it necessary to remove them before carrying out such studies. Thus, additional steps of proteolytic cleavage and subsequent removal of the proteolytic enzyme and fusion tag are applied to produce tag-free target proteins of interest. Widely used and commercially available proteases for the specific cleavages are thrombin, factor Xa, enterokinase, TEV protease, and precision protease [12]. Due to their lower specificity and instability of the target proteins, undesired cleavages have been observed outside the expected cleavage site [13]. For example, thrombin cannot differentiate between Ser and Cys in their recognition [14]. Moreover, these commercial enzymes might not be cost-effective when large-scale protein production is required, restricting their industrial scale applications in biotechnology. An alternative might be the use of thiol-inducible self-cleavable intein tags, which do not require additional proteases [15]. Instead, it utilizes an autocatalytic self-cleavage reaction induced by thiol reagents to avoid this problem. However, premature cleavage has been observed, thereby reducing the purification efficiency, and they often require optimization of parameters such as expression temperature and junction sequences [16,17]. In addition, the reducing condition used for the cleavage might not be compatible with some target proteins bearing disulfide bridges.

Here we report a set of new vectors in which small ubiquitin-like modifier (SUMO or SUMO homologue, SMT3) from yeast is used as cleavage tag, in tandem with other fusion tags such as TRX, TF, MBP and GST for solubility and expression enhancement. These tandem fusion vectors utilize the high specificity of ubiquitin-like protease 1 (Ulp1), which recognizes the three-dimensional structure of SUMO domain and cleaves after di-glycine at the C-terminus [18]. We demonstrated soluble expression and purification of the

Abbreviations: HSQC, heteronuclear single quantum coherence NMR spectrum; RF cloning, restriction free cloning (cloning which is independent of restriction sites).

* Corresponding author.
E-mail address: hideo.iwai@helsinki.fi (H. Iwaï).

http://dx.doi.org/10.1016/j.pep.2015.08.019
1046-5928/

C-terminal domain (CTD) of TonB protein from three organisms, and a small lectin protein scytovirin for optimal choice of a tandem fusion vector.

2. Materials and methods

2.1. Construction of plasmids

The backbone plasmid used to construct the expression vectors was pHYRSF53 [19,20]. Four different fusion tags were inserted in frame upstream of the coding sequence of SMT3. Two PCR steps were necessary to add a hexa-histidine (H_6) tag at the N-terminus of the cloned fusion tag. In a first PCR step, the coding sequences of the fusion tags were amplified using synthetic oligonucleotides. In all cases the forward primer contained an overhang coding part of a H_6-tag. The full H_6-tag and an *Nco*I site for cloning were added in a second PCR reaction using the first PCR product as a template and the oligonucleotide HK683: 5′-TAC CATGGGCAGCAGCCATCATCATCATCACGG as a forward primer. The resulting amplicon containing the H_6-fusion tag was inserted into the vector pHYRSF53 using the restriction sites *Nco*I and *Spe*I to generate the tandem SUMO-fusion vectors. The backbone pHYRSF53 and the four vectors generated are shown in Fig. 2A. To obtain pLJSRSF3, the glutathione S-transferase (GST) coding region was amplified with the primers I399: 5′-CATCATCATCAT CACGGCTCCCCTATACTAGGTTATTG and I398: 5′-TTTACTAGTTTTTG GAGGATGGTCGCCACC using the vector pGEX-2TK (GE Healthcare) as a template and cloned into pHYRSF53 as described above. The vector pLJSRSF7 was generated in the same way, amplifying the gene of maltose binding protein (MBP) from the plasmid pTWIN-MBP1 (New England Biolabs) using the primers I397: 5′-CATCAT CATCATCACGGCAAAATCGAAGAAGGTAAAC and I395: 5′-AAACTAG TACCCGAATTAGTCTGCGCGTC. The plasmid pCARSF85 was created similarly, amplifying trigger factor (TF) directly from *E. coli* genomic DNA using the primers I09: 5′-ATCATCATCATCATCACGGT CAAGTTTCAGTTGAAACC and I08: 5′-AAACTAGTACCTCCACCCGCCT GCTGGTTCATCAGC. Finally, to generate the plasmid pCARSF63 thioredoxin (TRX) was also cloned directly from *E. coli* genomic DNA using the primers HK682: 5′-CATCATCATCATCACGGCAGCGA TAAAATTATTCACC and HK684: 5′-CCACTAGTTCCCGCCAGGT TAGCGTCGAGG, and inserted into pHYRSF53 after the second PCR reaction as described above. The plasmids pHYRSF53, pLJSRSF3, pLJSRSF7, pCARSF85 and pCARSF63 are deposited and available for academic researchers from Addgene (addgene.org) with the deposit numbers #64696, #64692, #64693, #64694 and #64695, respectively.

The construction of the plasmid pJDJRSF05 carrying SMT3 fused with the single chain *Npu*DnaE intein variant was described previously [21]. The linker region was shortened by one Gly residue compared to pJDJRSF05 by amplifying the sequence of an inactive variant of single chain *Npu*DnaE intein using the synthetic oligonucleotides HK202: 5′-GTGGATCCGGAGCTCTAAGCTATGAAACG and SK187: 5′-ATCAAGCTTAATTAGAAGCTATGAAGCC. The PCR product was digested with *Bam*HI and *Hin*dIII and cloned into the vector pHYRSF53 to generate the plasmid pJDJRSF04 (Fig. 3). Likewise, to lengthen the linker region by one Gly residue compared to pJDJRSF05, a similar approach was carried out using the synthetic oligonucleotides HK204: 5′-GTGGATCCGGAGGAGGAGCTCTAAGC TATGAAACG and SK187. The resulting plasmid was pJDJRSF06. These three constructs were used to assess the $Ulp1_{403-621}$ activity.

The expression vectors constructed allow the possibility to add another two features downstream of the protein of interest (POI): the N-term of the *Npu*DnaE split intein, and a chitin binding domain. In this work we inserted the POI into the sites *Bam*HI and *Hin*dIII to avoid modifications in the C-terminus (Fig. 2A). Four different proteins were used in this study: the C-terminal domain (CTD) of TonB from three different organisms: *E. coli* (*Ec*TonB), *Pseudomonas aeruginosa* PAO1 (*Ps*TonB), and *Helicobacter pylori* (*Hp*TonB), and the antiviral lectin scytovirin (SVN) from *Scytonema varium*. The coding sequences of TonB variants were amplified by PCR from purified genomic DNA using synthetic oligonucleotides. The fragment containing the last 89 residues of *Ec*TonB (*Ec*TonB-89) was cloned using I470: 5′-AAGGATCCGGACCACGCGCAT TAAGCCG and SK009: 5′-TACAAGCTTACTGAATTTCGGTGGTGCCG. The amplified PCR fragment was *Bam*HI-*Hin*dIII digested and inserted into the *Bam*HI-*Hin*dIII digested expression vectors pHYRSF53, pLJSRSF3, pLJSRSF7, pCARSF85, pCARSF63 to generate the plasmids pFGRSF01, pFGRSF02, pFGRSF03, pFGRSF04, pFGRSF05, respectively. For *Ps*TonB two versions with different lengths were generated: the last 77 residues of TonB (*Ps*TonB-77) and the last 96 residues (*Ps*TonB-96). To clone the last 77 residues (*Ps*TonB-77) we used the synthetic oligonucleotides HK062: 5′-A AGGATCCCGGATGGCCCAGGCGCGGCG and HK063: 5′-TACAAGCT

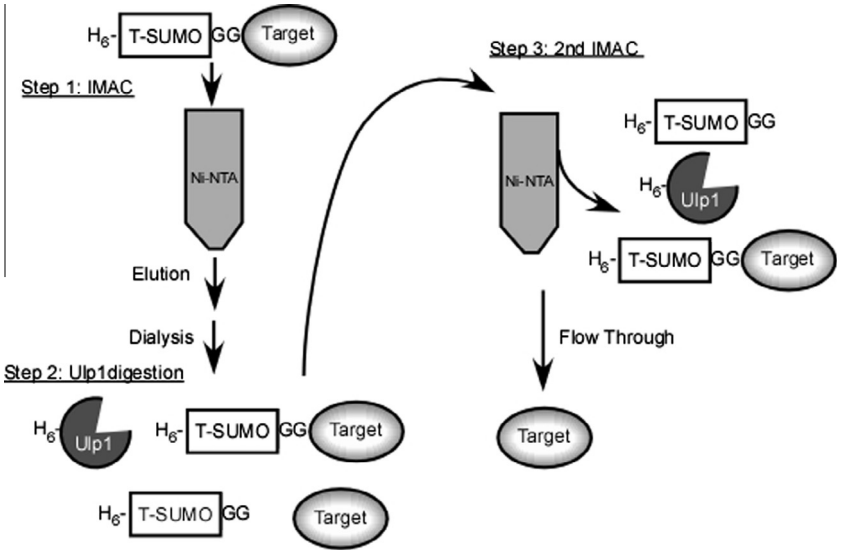

Fig. 1. Overview of the purification procedure of the target protein using the tandem SUMO vectors. T-SUMO: tandem fusion bearing yeast SMT3 as protease recognition site. Ulp1: protease domain (residues 403–621) of the *Saccharomyces cerevisiae* ubiquitin-like-specific protease 1.

Please cite this article in press as: F. Guerrero et al., Tandem SUMO fusion vectors for improving soluble protein expression and purification, Protein Expr. Purif. (2015), http://dx.doi.org/10.1016/j.pep.2015.08.019