

Appl Microbiol Biotechnol (2013) 97:6779–6791  
DOI 10.1007/s00253-012-4559-1

BIOTECHNOLOGICALLY RELEVANT ENZYMES AND PROTEINS

# The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology

Sofia J. Costa · André Almeida · António Castro ·  
Lucília Domingues · Hüseyin Besir

Received: 1 August 2012 / Revised: 24 October 2012 / Accepted: 28 October 2012 / Published online: 20 November 2012  
© Springer-Verlag Berlin Heidelberg 2012

**Abstract** The *Escherichia coli* host system is an advantageous choice for simple and inexpensive recombinant protein production but it still presents bottlenecks at expressing soluble proteins from other organisms. Several efforts have been taken to overcome *E. coli* limitations, including the use of fusion partners that improve protein expression and solubility. New fusion technologies are emerging to

complement the traditional solutions. This work evaluates two novel fusion partners, the Fh8 tag (8 kDa) and the H tag (1 kDa), as solubility enhancing tags in *E. coli* and their comparison to commonly used fusion partners. A broad range comparison was conducted in a small-scale screening and subsequently scaled-up. Six difficult-to-express target proteins (RVS167, SPO14, YPK1, YPK2, Frutalin and CP12) were fused to eight fusion tags (His, Trx, GST, MBP, NusA, SUMO, H and Fh8). The resulting protein expression and solubility levels were evaluated by sodium dodecyl sulfate polyacrylamide gel electrophoresis before and after protein purification and after tag removal. The Fh8 partner improved protein expression and solubility as the well-known Trx, NusA or MBP fusion partners. The H partner did not function as a solubility tag. Cleaved proteins from Fh8 fusions were soluble and obtained in similar or higher amounts than proteins from the cleavage of other partners as Trx, NusA or MBP. The Fh8 fusion tag therefore acts as an effective solubility enhancer, and its low molecular weight potentially gives it an advantage over larger solubility tags by offering a more reliable assessment of the target protein solubility when expressed as a fusion protein.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00253-012-4559-1) contains supplementary material, which is available to authorized users.

S. J. Costa · L. Domingues (✉)  
IBB - Institute for Biotechnology and Bioengineering,  
Centre of Biological Engineering, Universidade do Minho,  
Campus de Gualtar,  
4710-057 Braga, Portugal  
e-mail: [luciliad@deb.uminho.pt](mailto:luciliad@deb.uminho.pt)

S. J. Costa · A. Castro  
Instituto Nacional de Saúde  
Dr. Ricardo Jorge (INSARJ),  
Porto, Portugal

A. Almeida  
Hitag Biotechnology, Lda.,  
Biocant-Parque Tecnológico de Cantanhede,  
Coimbra, Portugal

A. Almeida  
Instituto de Ciências e Tecnologias Agrárias e  
Agro-Alimentares (ICETA), Universidade do Porto,  
Porto, Portugal

H. Besir  
Protein Expression and Purification Core Facility,  
European Molecular Biology Laboratory,  
Heidelberg, Germany

**Keywords** *Escherichia coli* · Fusion protein · Fh8 fusion tag · Traditionally used fusion tags · Protein solubility · Tag removal

## Introduction

The production of soluble and functional protein in *Escherichia coli* is still a major challenge in biotechnology

research. In spite of its fast growth, low cost, high productivity, and extensive genetic characterization, *E. coli* occasionally still suffers from low expression and/or low solubility of target proteins (Demain and Vaishnav 2009; Makino et al. 2011; Peti and Page 2007; Terpe 2006). Several efforts have been exploited to prevent recombinant protein aggregation and to improve its soluble production by the use of different promoters, expression strains and induction conditions, co-expression of chaperones, and soluble fusion partners (Berrow et al. 2006; Makino et al. 2011; Pacheco et al. 2012; Sorensen and Mortensen 2005; Studier 2005; Vernet et al. 2011). The fusion of a highly soluble carrier to recombinant proteins has been generally used to improve protein solubility and expression in *E. coli* (Esposito and Chatterjee 2006; Ohana et al. 2009; Terpe 2003; Waugh 2005)), although success is not yet always guaranteed. The Trx (LaVallie et al. 2000), GST (Smith 2000; Smith and Johnson 1988), MBP (Kapust and Waugh 1999; Sachdev and Chirgwin 2000), and NusA (Davis et al. 1999; De Marco et al. 2004) fusion partners are commonly employed as solubility enhancing carriers, but when producing a recombinant protein for structural and functional applications, these fusion partners must often be removed. The removal of fusion partners is usually made by specific protease sites included between the fusion tag and the target protein. However, after cleavage of the soluble fusion partner, precipitation of the target proteins can occur. Here, a major bottleneck appears as the target protein solubility can dramatically change in the presence and absence of the fusion partner. Meanwhile, new fusion solutions are constantly emerging and complementing the other fusion partners, as for instance, SUMO fusion technology (Marblestone et al. 2006; Malakhov et al. 2004). A novel fusion system (Hitag<sup>®</sup>) is presented in this work: the Fh8 and H partners. The Fh8 fusion partner is an 8-kDa calcium-binding recombinant protein (GenBank ID AF213970) extracted from the parasite *Fasciola hepatica*, and it has been previously used on the diagnosis of parasite infections (Silva et al. 2004). The recombinant Fh8 was also studied before by directed mutagenesis, in which its N-terminal sequence revealed to be important for Fh8 stability and production in *E. coli* (not published). From this analysis, a new fusion partner was suggested: the H tag that corresponds to the first eleven aminoacids of the Fh8 N-terminus, resulting in a molecular weight of 1 kDa. In this work, both Fh8 and H fusion partners are explored as solubility enhancing partners and compared to the commonly used fusion partners Trx, GST, MBP, NusA, and SUMO. The study conducted here does not only evaluate the novel fusion system effect on protein solubility but also the behavior of target proteins after Fh8 and H tags removal in comparison to the other fusion partners. Six difficult-to-express target proteins in *E. coli* were fused to eight fusion tags and the resulting solubility compared in a broad screening before and after tag removal.

## Materials and methods

### General

In this work, all the cloning PCRs used the Phusion High-Fidelity DNA Polymerase (New England Biolabs) with an annealing temperature of 55 °C, according to the manufacturer's instructions. The colony PCRs were conducted using the in-house DNA Taq Polymerase with an annealing temperature of 55 °C and with the T7 forward and reverse universal primers. Plasmid DNA extractions were performed using the QIAGEN kits for maxi- and minipreps and the QIAquick DNA gel extraction kit or QIAquick PCR purification kit (Qiagen) were used for DNA purification. The restriction enzymes used in this work were from New England Biolabs. All the DNA ligations were carried out with the Rapid DNA Ligation kit (Roche). For plasmid maintenance and protein expression, different antibiotics were used depending on the strain and plasmid requirements. Antibiotic stock solutions were prepared, filtered through 0.2 µm, and stored at -20 °C in the following concentrations: kanamycin, 30 mgmL<sup>-1</sup>; carbenicillin, 100 mgmL<sup>-1</sup>; and chloramphenicol, 10 mgmL<sup>-1</sup>.

### Construction of the pETMFh8 and pETMH vectors

The insertion of the Fh8 tag with the TEV cleavage site into the pETM10 vector (Table 1) was carried out by DNA ligation of the *NcoI*-*KpnI* digested plasmid and *PciI*-*KpnI* digested *fh8* PCR product. The pETMH fusion vector was obtained from the pETM11 backbone (Table 1), performing three PCRs: The PCR-I inserted part of the H tag (the first 28 nucleotides) after the His<sub>6</sub> tag sequence from the pETM11 plasmid. The PCR-II inserted part of the H tag (the final 28 nucleotides) before the TEV cleavage site of the pETM11 plasmid. As the H tag sequence has only 33 basepairs (bp), specific primers were designed for the PCR I and II in order to have 23 nucleotides of the H tag sequence matching in both PCRs. The universal primers T7 forward and reverse were used in the PCR-III for amplification of the His<sub>6</sub> tag/H tag/TEV site sequence to be cloned into the pETM11 plasmid. The purified PCR-III product and the pETM11 plasmid were digested with *XbaI* and *XhoI* restriction enzymes, and the final pETMH vector was obtained by ligation of the digested DNAs. *E. coli* DH5α competent cells were transformed with the constructed pETMFh8 and pETMH plasmids, and the obtained clones were analyzed by colony PCR. The novel pETMFh8 and pETMH fusion vectors were confirmed by sequencing with both T7 forward and reverse universal primers.

### Cloning of the target genes into pETM vectors

The target genes *ypk1\_frag*, *rvs167*, *spo14\_frag*, and *ypk2* used in this work (Table 2) are synthetic genes previously

**Table 1** Construction of the pETMFh8 and pETMH expression vectors

Backbone	Vector	PCR	Primers	Sequence (5'→3')	Comments
pETM 10	pETMFh8	I	Fh8-FWD	<u>TCTATTACATGTCCCCTAGTGTCAAGA</u> <u>GGTTGAAAAAC</u>	In bold is the <i>PciI</i> restriction enzyme sequence and underlined is the initial part of the Fh8 tag sequence
			Fh8-RV	<u>AATAGAGGTACCGGAACCCATGGAGCC</u> <u>CTGAAAATAAAGATTCTCTGACAAAATCG</u> <u>AAACGAG</u>	In bold is the <i>KpnI</i> restriction enzyme sequence and in bold underlined is the <i>NcoI</i> restriction enzyme sequence. In italic bold is the TEV recognition cleavage site and underlined is the final part of the Fh8 tag sequence
pETM 11	pETMH	I	T7-FWD	<u>TAATACGACTCACTATAGGG</u>	Underlined are the first 28 nucleotides of the H tag sequence. In bold underlined are the 23 common nucleotides used in both I and II PCRs. In italic bold is the His <sub>6</sub> tag sequence of the pETM11 plasmid
			Htag-RV	<u><b><i>GTTTTTCAACCTCTTGAACACTAGGCAT</i></b></u> <u><b><i>GTGATGGTATGGTGATGTTTC</i></b></u>	
		II	Htag-FWD	<u><b><i>TAGTGTTCAAGAGGTTGAAAACTCCTT</i></b></u> <u><b><i>GAGAATCTTTATTTTCAGGGC</i></b></u>	Underlined are the final 28 nucleotides of the H tag sequence. In bold underlined are the 23 common nucleotides used in both I and II PCRs. In italic bold is the TEV recognition cleavage site of the pETM11 plasmid
			T7-RV	<u>GCTAGTTATTGCTCAGCGG</u>	

optimized for *E. coli* expression. All these genes were inserted into the pETM vectors (Table 3) using the *NcoI/BamHI-XhoI* restriction sites, presented at the beginning of each gene sequence and after the stop codon, respectively. The *frutalin* and *cp12* genes (Table 2) were modified by PCR to have the *NcoI/BamHI-XhoI* restriction sites at the same positions as the other target genes. A similar primer designing was used to amplify both *frutalin* and *cp12* sequences as follows: primer forward 5'-TCTATTCC ATGGGATCC-22 initial nt of the target gene -3' with the *NcoI* restriction site in bold, the *BamHI* restriction site in italic and the first 22 nucleotides of each gene underlined, and primer reverse 5'-AATAGACTCGAG-22 final nt of the target gene-3' with the *XhoI* restriction site in bold and the final 22 nucleotides of each gene underlined. After ligation of the digested PCR products and plasmids, *E. coli* DH5 $\alpha$  competent cells were transformed, and the resulting clones were analyzed and confirmed as previously mentioned.

#### Expression strains and culture conditions

Four different expression strains were evaluated in this work: the *E. coli* BL21 (DE3) Codon Plus-RIL, the *E. coli* Rosetta (DE3) cells, the *E. coli* Tuner (DE3) strain, and the *E. coli* SoluBL21 strain. Competent cells of the four different strains were prepared and transformed with the constructed fusion vectors. The resulting clones were confirmed by colony PCR, and one positive clone was selected for the expression trials. All the cells were grown in LB media with the appropriate antibiotics diluted with a factor of 1,000 (see Table 3 for selection and maintenance of pETM vectors). In this work, the degradation-resistant carbenicillin was used instead of

ampicillin. Precultures were grown overnight (o/n) at 37 °C and a dilution factor of 100 was used for inoculation of all cultures (usually corresponding to a starting OD<sub>600nm</sub> of 0.02). Cultures were performed in parallel, using 10 mL of culture media in 24 deep-well plates (25 mL capacity per well; DIA Nielsen GmbH&Co. KG, Germany) for small-scale screenings and 500 mL of culture media in 2-L flasks for the scale-up. Cultures were grown at 37 °C and 200 rpm to a final OD<sub>600nm</sub> of 0.4–0.6 before induction. In the small-scale study, two plates with the same strain and fusion proteins were used to test different induction conditions: isopropyl- $\beta$ -D-1-thiogalactopyranoside at 0.5 mM, 28 °C and 3 h (first plate) or at 0.2 mM, 18 °C and o/n (second plate). Each 10 mL culture was divided in two 5-mL cultures. All cells were harvested for 25 min, at 4 °C and 4,500 rpm. Cell pellets from 500 mL cultures were washed once with 1 $\times$  phosphate-buffered saline and collected again by centrifugation. Bacterial pellets were flash frozen in liquid nitrogen and stored at -20 °C.

#### Cell lysis

In the small-scale screening, two sonication protocols were tested using an eight-microtip sonicator or a single-tip sonicator (G. Heinemann, Germany). For the eight-microtip lysis, cells from a 5-mL culture pellet were resuspended in 1 mL of lysis buffer [50 mM Tris pH8.0, 250 mM NaCl, 20 mM imidazole buffer supplemented with 1 $\times$  complete free EDTA protease inhibitor (Roche), 5 mM MgCl<sub>2</sub> (Sigma), 5  $\mu$ g mL<sup>-1</sup> DNase (Sigma), and 1 mg mL<sup>-1</sup> lysozyme (Sigma)] and transferred to a 96-deep-well plate (2 mL capacity per well) and incubated at room temperature for 10 min. The plate was placed on ice, and cells were

**Table 2** Features and properties of target genes used in this study

Synthetic genes	GenBank/Uniprot	Organism	Localization	Function or application	Gene size (nt)	Protein size (aa)	MW (Da)	pI	Cysteines (%)	GRAVY
<i>frutalin</i>	Jacalin-related lectin: L03795.1 Q38720	<i>Artocarpus incisa</i>	Plant seeds	Biomarker of cancer cells [27]	477	158	17,194	8.73	0	-0.089
<i>cp12fragment</i>	XM_625821.1 Q5CR31	<i>Cryptosporidium parvum</i>	Oocyst surface	Involved in the host-control interaction [25]	216	71	8,414.2	4.72	0	-1.104
<i>ypk1fragment</i>	NM_001179692.1 P12688	<i>Saccharomyces cerevisiae</i>	Cell membrane and cytoplasm	Proliferation of yeast cells; involved in signaling pathways (Uniprot)	964	312	35,629	6.54	1.6	-0.224
<i>rvs167</i>	NM_001180696.1 P39743	<i>Saccharomyces cerevisiae</i>	Cytoplasm and cytoskeleton	Formation of endocytic vesicles and cytoskeletal reorganization (Uniprot)	1,516	496	54,134	5.77	0.2	-0.470
<i>spo14fragment</i>	NM_001179821.1 P36126	<i>Saccharomyces cerevisiae</i>	Endosome, nucleus and prospore membrane	Meiosis and spore formation (Uniprot)	1,192	388	44,828	8.56	0.5	-0.579
<i>ypk2</i>	NM_001182604.1 P18961	<i>Saccharomyces cerevisiae</i>	Nucleus and cytoplasm	Proliferation of yeast cells; involved in signaling pathways (Uniprot)	2,104	692	78,111	7.33	0.9	-0.427

*nt* nucleotides, *aa* amino acids, *MW* molecular weight, *Da* Dalton, *pI* isoelectric point, *GRAVY* grand average of hydropathicity

further lysed by sonication. The lysis efficiency was improved by adding 400  $\mu$ L of 0.5 mm beads to each well. The 96-deep-well plates were then centrifuged at 4,000 rpm, 4  $^{\circ}$ C for 45 min, and the supernatant fraction was collected to a new 96-deep-well plate. For the single-tip lysis, 5-mL cell pellets were resuspended in 1 mL of lysis buffer and transferred to a 2-mL tube containing 0.5 mm beads. After 10 min incubation at room temperature, bacterial cells were lysed by sonication, and 2-mL tubes were centrifuged at 13,000 rpm at 4  $^{\circ}$ C for 25 min. The supernatant fractions were transferred to new 2-mL tubes. In the scale-up experiments, cell pellets were thawed and resuspended in 10 mL of lysis buffer. After incubation at room temperature for 10 min, cells were sonicated (Branson Sonifier 250, G. Heinemann, Germany), and the soluble fraction was removed from the insoluble cell debris by ultracentrifugation at 40,000 rpm, 4  $^{\circ}$ C for 30 min. In all the experiments, aliquots of total lysates and supernatant samples were taken and prepared to be analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

#### Protein purification and tag removal

The selected fusion proteins were purified by nickel affinity chromatography using a spin protocol with Ni-NTA slurry (Qiagen) for the small-scale screening or a semi-automated system, in which 100  $\mu$ L prepacked Ni-NTA superflow columns (Robot Columns, Atoll GmbH, Germany) were set on a 96-well plate matrix, for the scale-up experiments. Both purifications were conducted according to manufacturer's instructions, using 50 mM Tris pH 8.0, 250 mM NaCl with 20 mM imidazole as running and washing buffer, and with 300 mM imidazole for the elution buffer. The collected fractions were analyzed by SDS-PAGE, and the total protein content of robot eluted samples was also estimated by Bradford assay.

For tag removal, the selected and purified proteins were digested with the TEV protease. After digestion, aliquots were taken, and samples were centrifuged at 13,000 rpm for 10 min at 4  $^{\circ}$ C. The cleaved proteins were then purified from the fusion tags and proteases by nickel affinity chromatography using the same above-mentioned protocol. The collected samples were prepared to be analyzed by SDS-PAGE.

#### Protein expression and solubility evaluation

The expression and solubility evaluation was conducted using a score of 0, 1, 2, and 3, based on the expected soluble protein production yields obtained after the purification of 1-L cultures (Berrow et al. 2006; Bird 2011). The score 0 corresponds to no expression/soluble protein, a score of 1 was given to the soluble expressions that are expected to yield <0.5 mg, the score of 2 indicates a soluble expression yield between 0.5 and

**Table 3** Features of expression vectors and properties of the tags used in this work

Expression vectors (pETM)	Fusion protein	Tag size (aa)	pI	MW tag (Da)	GRAVY	Promoter	Selection	Protease cleavage site
11 (His)	MK- <i>His<sub>6</sub></i> -TEV(ENLYFQ↓G)AMGS-Target	29	6.24	3,420	-1.09	T7/lac	kan	TEV
20 (Trx)	M- <i>Trx</i> - <i>His<sub>6</sub></i> -TEV(ENLYFQ↓G)AMGS-Target	136	5.66	14,673	-0.189	T7/lac	amp	TEV
30 (GST)	MK- <i>His<sub>6</sub></i> - <i>GST</i> -TEV(ENLYFQ↓G)AMGS-Target	254	6.31	29,264	-0.446	T7/lac	kan	TEV
41 (MBP)	MK- <i>His<sub>6</sub></i> - <i>MBP</i> -TEV(ENLYFQ↓G)AMGS-Target	404	5.49	44,549	-0.462	T7/lac	kan	TEV
60 (NusA)	M- <i>NusA</i> - <i>His<sub>6</sub></i> -TEV(ENLYFQ↓G)AMGS-Target	519	4.63	57,383	-0.306	T7/lac	kan	TEV
SUMO	MK- <i>His<sub>6</sub></i> -TEV(ENLYFQG)- <i>SUMO</i> -TGGS-Target	108	6.02	12,453	-0.882	T7/lac	kan	SUMO
H	MK- <i>His<sub>6</sub></i> - <i>Htag</i> -TEV(ENLYFQ↓G)AMGS-Target	30	6.56	3,553	-0.773	T7/lac	kan	TEV
Fh8	MK- <i>His<sub>6</sub></i> - <i>Fh8tag</i> -TEV(ENLYFQ↓G)SMGS-Target	90	6.43	10,146	-0.780	T7/lac	kan	TEV

aa aminoacids, pI isoelectric point, MW molecular weight, Da Dalton, GRAVY grand average of hydrophaticity

5 mg, and a score of 3 specifies the soluble expressions with an expected yield >5 mg. The soluble expression results were estimated from the SDS-PAGE analysis of

the eluted fractions from nickel column and also from the respective supernatant fractions, according to the following equations:

$$\text{Protein soluble expression} = \frac{\text{number of scores 1, 2, and 3 for the tested protein}}{\text{number of fusion tags tested}} \quad (1)$$

$$\text{Tag soluble expression} = \frac{\text{number of soluble proteins fused to the tag with scores 1, 2, and 3}}{\text{number of total proteins tested}} \quad (2)$$

$$\text{Tag soluble score 2 + 3} = \frac{\text{number of soluble proteins fused to the tag with scores 2 and 3}}{\text{number of total proteins tested}} \quad (3)$$

The total expression results were estimated from the SDS-PAGE analysis of the total lysate fractions using the

same scores 0–3, but here for the expected expression yields, as follows:

$$\text{Tag total expression} = \frac{\text{number of expressed proteins with scores 1, 2, and 3}}{\text{number of total proteins tested}} \quad (4)$$

## Results

### The novel pETMFh8 and pETMH fusion vectors

The fusion tags in this study, Fh8 and H tags, were inserted into the pETM vector series (Dummler et al. 2005) resulting into two novel expression vectors: the pETMFh8tag and pETMHtag (see Online Resource 1). The pETMFh8tag vector was constructed using the pETM10 backbone and the

pETMHtag vector used the pETM11 backbone (Table 1). Both new vectors share common features with the rest of the pETM vectors used in this work (Table 3): They have similar multiple cloning sites (MCS) suitable for direct subcloning of the target genes; two *His<sub>6</sub>* tag sequences, one placed at the N-terminal of the fusion partner and the other after the MCS; and a TEV protease cleavage site between the fusion partner and the MCS. The novel fusion plasmids are kanamycin resistant. In this work, only the N-terminal *His<sub>6</sub>* tag was used for protein



purification. In the pETM series, the Fh8 and H tags have only 10 and 3.6 kDa, respectively (Table 3). Both new fusion tags have lower molecular weights than the studied fusion tags, even compared to the low molecular weight tags such as SUMO (12.5 kDa) and Trx (14.7 kDa). The analysis of fusion tags features (Table 3) shows that Fh8 and H tags have identical isoelectric points, which are also similar to the GST or SUMO values. The Grand average of hydropathicity (GRAVY) shows a predominantly hydrophilic nature of the fusion tags (all the GRAVY values are negative). The Fh8 and H tags along with SUMO and His tags present the highest hydrophilic character.

#### Cloning of target proteins into the pETM expression vectors

The synthetic genes used in this work (Table 2) were cloned in parallel into the pETM vectors using the same restriction enzymes for all the constructions: All the pETM vectors received *NcoI*–*XhoI* digested target genes with the exception for pETMSUMO, which received *Bam*HI–*XhoI* digested inserts. This strategy allowed a rapid and easy cloning procedure and a successful cloning rate of 98 %. From the 48 fusion genes to be constructed, only the cloning of the *ypk1\_frag* into the pETMhtag vector was not successful, as confirmed by sequencing at the GATC (Germany). The six target proteins used in this work were selected regarding their previous difficulty of expression in soluble form in *E. coli*. Four of these proteins (two full length, RVS167 and YPK2; and two truncated proteins, SPO14 and YPK1) are found in the yeast *Saccharomyces cerevisiae* and were chosen due to the little/absent soluble expression in *E. coli* obtained when formerly fused to human SUMO3 tag (data not shown). The other two target proteins [the truncated CP12 derived from the *Cryptosporidium parvum* parasite (Yao et al. 2007) and the lectin Frutalin from the plant seed *Artocarpus incisa* (Oliveira et al. 2009, 2011)] were previously expressed as full lengths in *E. coli* presenting poor solubility. The selected target proteins have different locations and functions or applications, and they also differ in size: the smallest target protein used in this work (a truncated form of the CP12 protein lacking its transmembrane region) has only 71 amino acids, and the largest one, YPK2, has approximately 700 amino acids. All the target proteins have low cysteine content, and they range from a slight to moderate hydrophilic nature (Table 2).

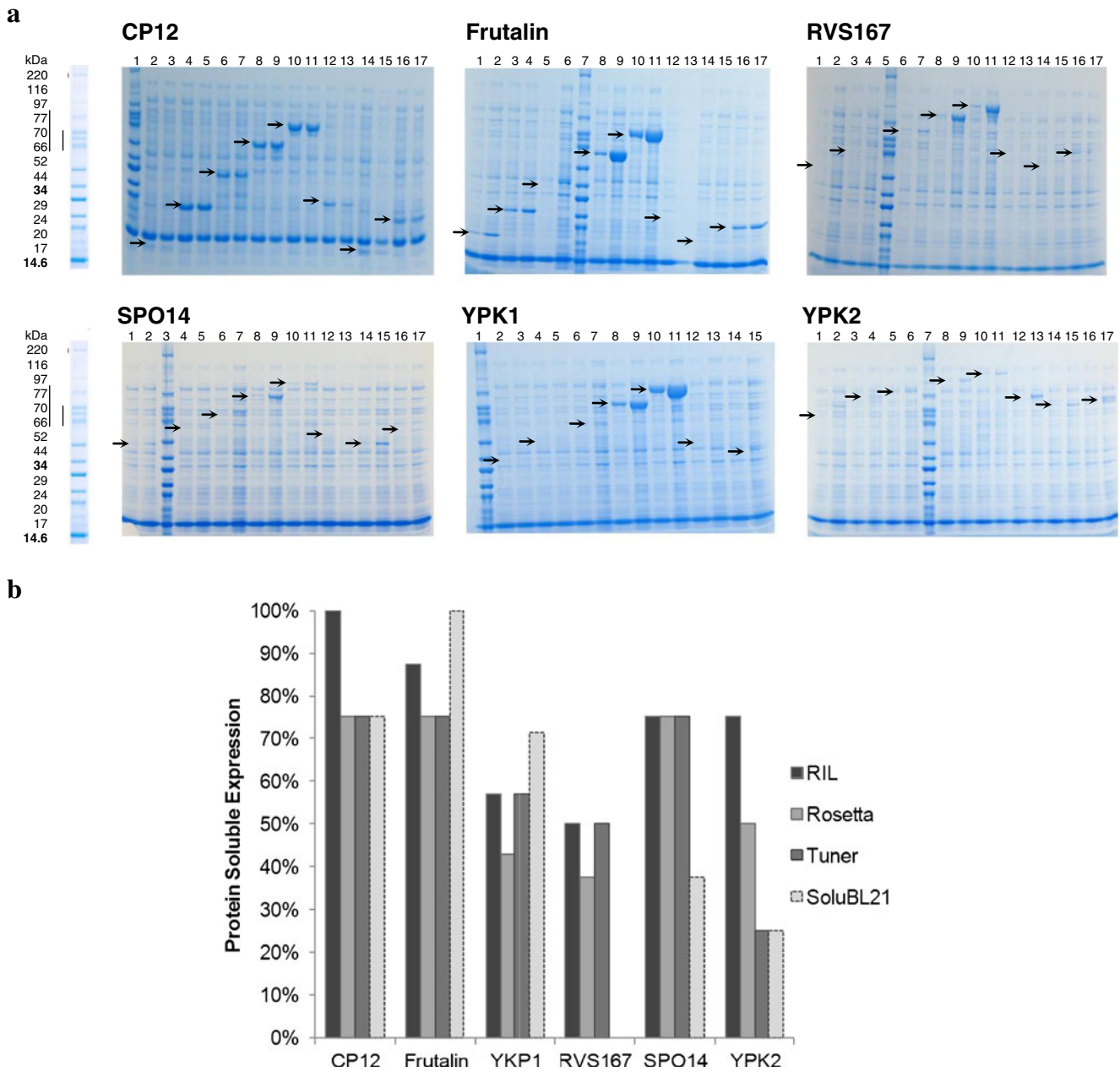
#### Small-scale screening: selection of the expression strain and culture conditions

Figure 1 presents the small-scale comparison of the soluble expression for the different strains, fusion tags, and target proteins. Figure 1a presents the SDS-PAGE analysis of supernatant samples from the selected strains, using the two

induction conditions tested. Figure 1b result from the analysis of scores 1–3 in supernatant fractions obtained from the eight-micropipette sonication protocol, using Eq. 1. For all the four strains, the number of soluble expressed proteins using the Fh8 tag was similar to the number obtained with the MBP or NusA tags, and higher than the number obtained using the Trx and GST tags. Proteins fused to the H tag and His<sub>6</sub> tag presented the lowest soluble expression along with proteins fused to SUMO, whose insolubility was previously evaluated for four of the six targets. The solubility of SUMO–Frutalin fusion was difficult to evaluate, as observed by the small protein content in lanes 12 and 13 of Fig. 1a. SUMO-CP12 and SUMO-YPK2 fusion proteins presented amounts of soluble protein similar to those of Fh8-CP12 and Fh8-YPK2, respectively, as observed in Fig. 1a. CP12 and Frutalin were soluble expressed within the highest number of fusion tags, and this result was observed among all the four strains used in this work (Fig. 1b). This small-scale screening resulted in the following rank of the fusion tags for protein solubility: MBP>NusA>Fh8~Trx>GST~H>His. According to the obtained results in Fig. 1, the *E. coli* BL21 (DE3) Codon Plus-RIL strain was established for the expression of the fusions with CP12, RVS167, and YPK2 target proteins, and the *E. coli* Rosetta (DE3) strain was selected for the expression of the fusions with Frutalin and truncated SPO14. Finally, fusions with the truncated YPK1 protein were later expressed using the *E. coli* Tuner (DE3) strain. The *E. coli* SoluBL21 strain was not used further in this study. Regarding the two different culture conditions tested, the overnight induction condition presented the highest level of soluble expression for five of the six target proteins. The 3-h induction condition was only the best option for the fusions with CP12 target protein. All the soluble expressed fusion proteins presented a molecular weight identical to the expected size (see Online Resource 2). SUMO fusions migrated on SDS-PAGE with a molecular weight of 3 kDa higher than the expected, which is also known from previous publications (Marblestone et al. 2006).

#### Small-scale expression and solubility results: validation

For each target protein, four to six tags were selected to continue with the small-scale screening and to validate the initial solubility results. Table 4 summarizes the selected fusion proteins for the direct comparison of commonly used fusion tags and Fh8 tag. Figure 2 presents the SDS-PAGE analysis of the total lysate and supernatant fractions of the selected fusion proteins that were obtained from the bacterial lysis with a single micropipette sonication. These samples were also scored 0–3 (see Online Resource 3). As observed, the MBP, NusA, and Trx fusions resulted in higher total expression than the Fh8 fusions. The Fh8 fusions resulted in a total expression identical to the H and His fusions. To further validate the solubility screening, the supernatant



**Fig. 1** Comparison of fusion protein soluble expression using different *E. coli* strains: small-scale screening evaluation of the supernatant fractions from the cell lysis with the eight-tip sonicator. **a** SDS-PAGE of the supernatant samples from the two induction conditions tested. The 3-h induction samples were loaded aside with the o/n induction, according to the following fusion tags order: His–Trx–GST–MBP–NusA–SUMO–

H–Fh8tag. The YPK1 gel follows the same order except for the H tag (which is not available with this target protein). The in-house protein marker was used for this analysis. Arrows indicate the expected molecular weights for all fusion proteins following the fusion tags order loaded on the gels. **b** Soluble expression (scores 1–3) per target protein with different strains. Plotted values were estimated by Eq. 1

samples presented in Fig. 2 were processed by nickel affinity small-scale purification. Results from this analysis were evaluated by SDS-PAGE and also scored 0–3 (Fig. 3a, b, respectively). The MBP and NusA fusion proteins improved the solubility of a higher number of proteins than the Fh8 tag (groups I, II and III; Table 4). Proteins fused to MBP and NusA tags were, however, obtained in lower soluble amounts than those estimated before purification (see

Online Resource 3e). For the Fh8 fusion proteins, no differences were observed before (see Online Resource 3) and after purification. The same observation is valid for the fusions of the comparison groups IV, V, and VI of Table 4 (see Online Resource 3 for the scoring results before purification). SUMO and Fh8 tags presented identical soluble amounts for CP12 and YPK2 proteins (Table 4, group V). The Fh8 fusions presented higher solubility than the H

**Table 4** Selection of fusion proteins for small-scale screening—comparison groups

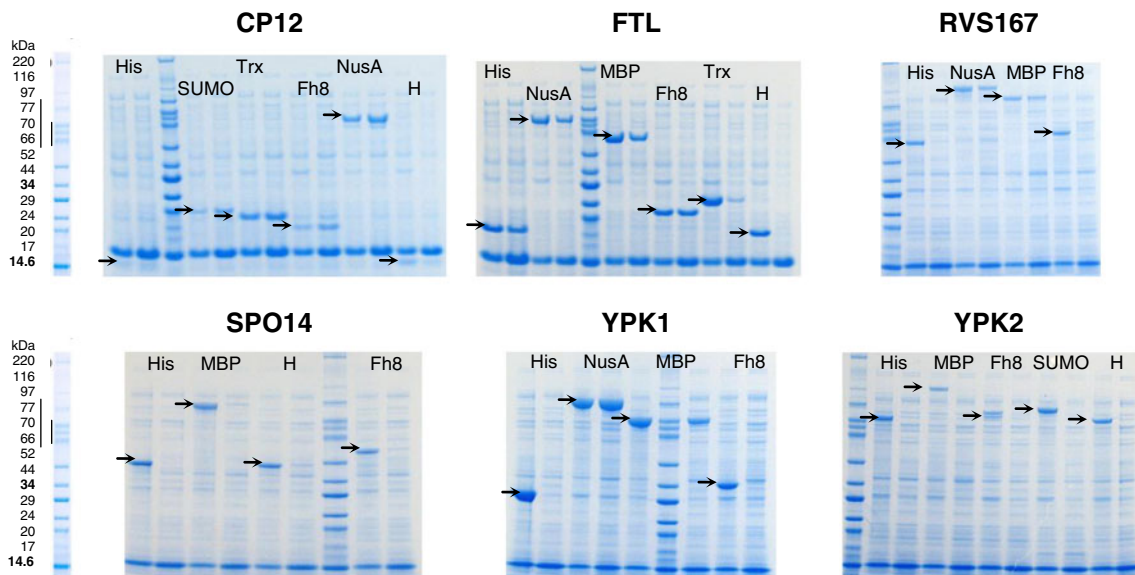
Comparison group no.	Selected fusion tags	Target proteins
I	His, MBP, Fh8	Frutalin, RVS167, SPO14, YPK1, YPK2
II	His, NusA, Fh8	CP12, Frutalin, RVS167, YPK1
III	His, MBP, NusA, Fh8	Frutalin, RVS167, YPK1
IV	His, Htag, Fh8	CP12, Frutalin, YPK2
V	His, SUMO, Fh8	CP12, YPK2
VI	His, Trx, NusA, H, Fh8	CP12, Frutalin

fusions. Interestingly, when directly compared using the CP12 and Frutalin proteins (Table 4, group VI), Fh8 fusion proteins achieved the highest solubility, identical to proteins fused with NusA tag. In general, the rank of the fusion tags obtained in this analysis for protein solubility after purification was identical to the previously observed, with the NusA and the MBP on the top of solubility tags, followed by the Fh8 and Trx tags. The H tag increased the *E. coli* protein expression identical to the Fh8 tag but did not improve protein solubility. The target protein SPO14 was not used further in this work as no soluble protein was detected in the small-scale expression and purification screenings.

Scale-up protein processing: evaluation of the fusion proteins solubility before and after tag removal

Figure 4 presents the SDS-PAGE results obtained for the scale-up screening. All the fusion proteins were successfully expressed in soluble form at the expected molecular weights (see Online Resource 2) in *E. coli* 500 mL cultures.

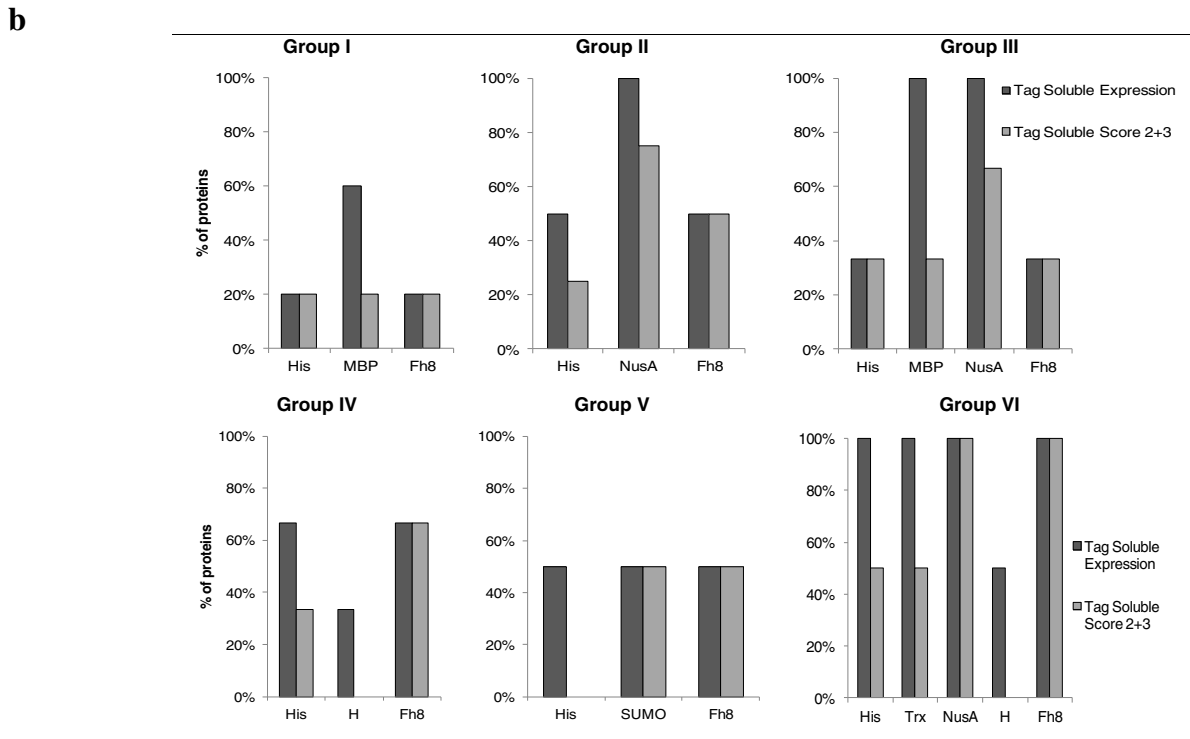
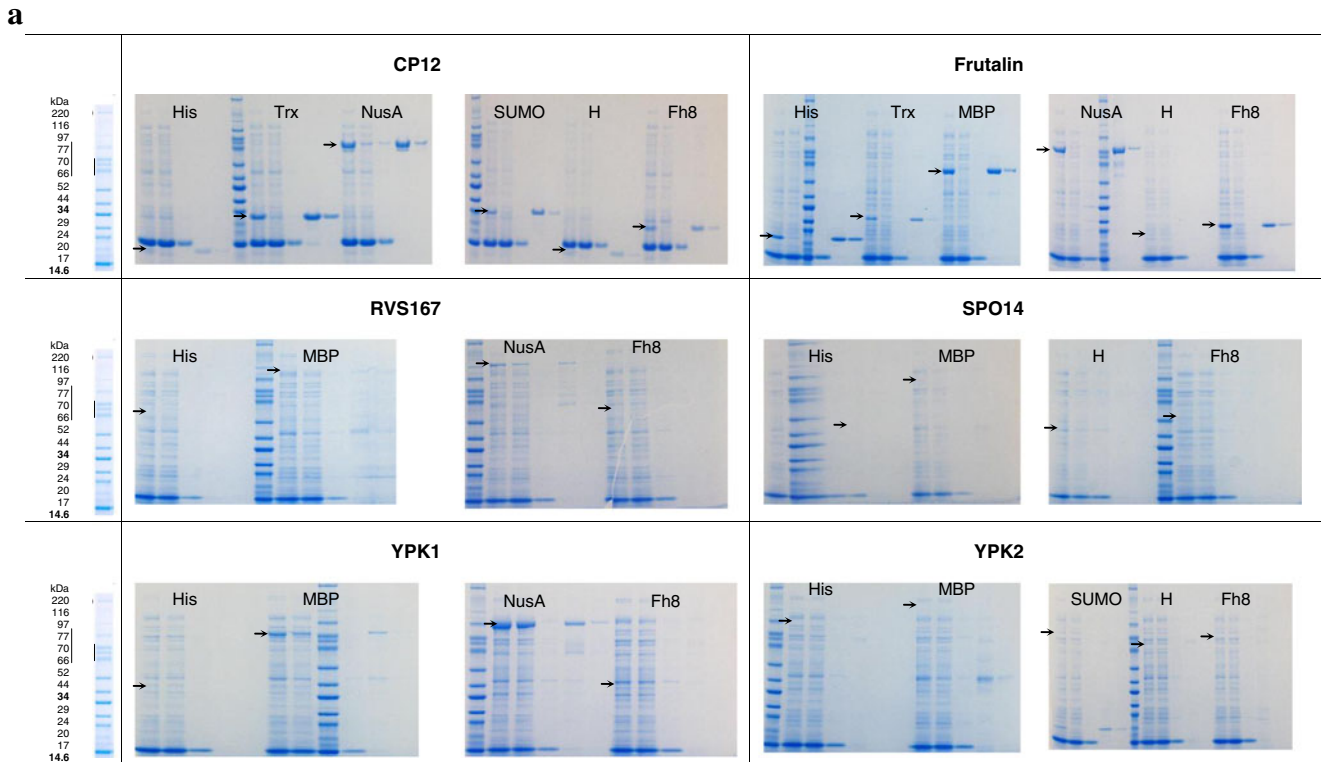
Bradford estimations were also conducted to support the SDS-PAGE analysis (see Online Resources 4 and 5). Making a wide comparison based on the group VI (Table 4), the Fh8 fusions resulted in similar soluble amounts as the Trx and NusA fusions, corroborating the small-scale screening results. After tag removal, the Fh8 fusions performed better than the NusA or MBP fusions, resulting in higher soluble amounts of the cleaved proteins. A case by case comparison established that NusA, Trx, and Fh8 fusions improved the Frutalin soluble expression following this order: NusA~Fh8>Trx>His>Htag (Fig. 4a). Interestingly, after tag removal, the cleaved and purified Frutalin (17 kDa) from the Fh8 and Trx fusions presented higher protein amounts than the Frutalin cleaved from the NusA fusion protein. A similar result was obtained for the expression of CP12 target protein with the different fusion tags (Fig. 4b). The Fh8-CP12 fusion protein achieved similar, but slightly lower, soluble amounts than the Trx and NusA fusions. After tag removal, the cleaved CP12 protein (8 kDa) from Fh8 fusion presented higher soluble amounts



**Fig. 2** SDS-PAGE analysis of the total lysate and supernatant samples from the selected fusion proteins in the second small-scale screening. Samples were obtained after cell lysis with a single tip sonication. Gels were loaded with each total lysate sample aside with the respective

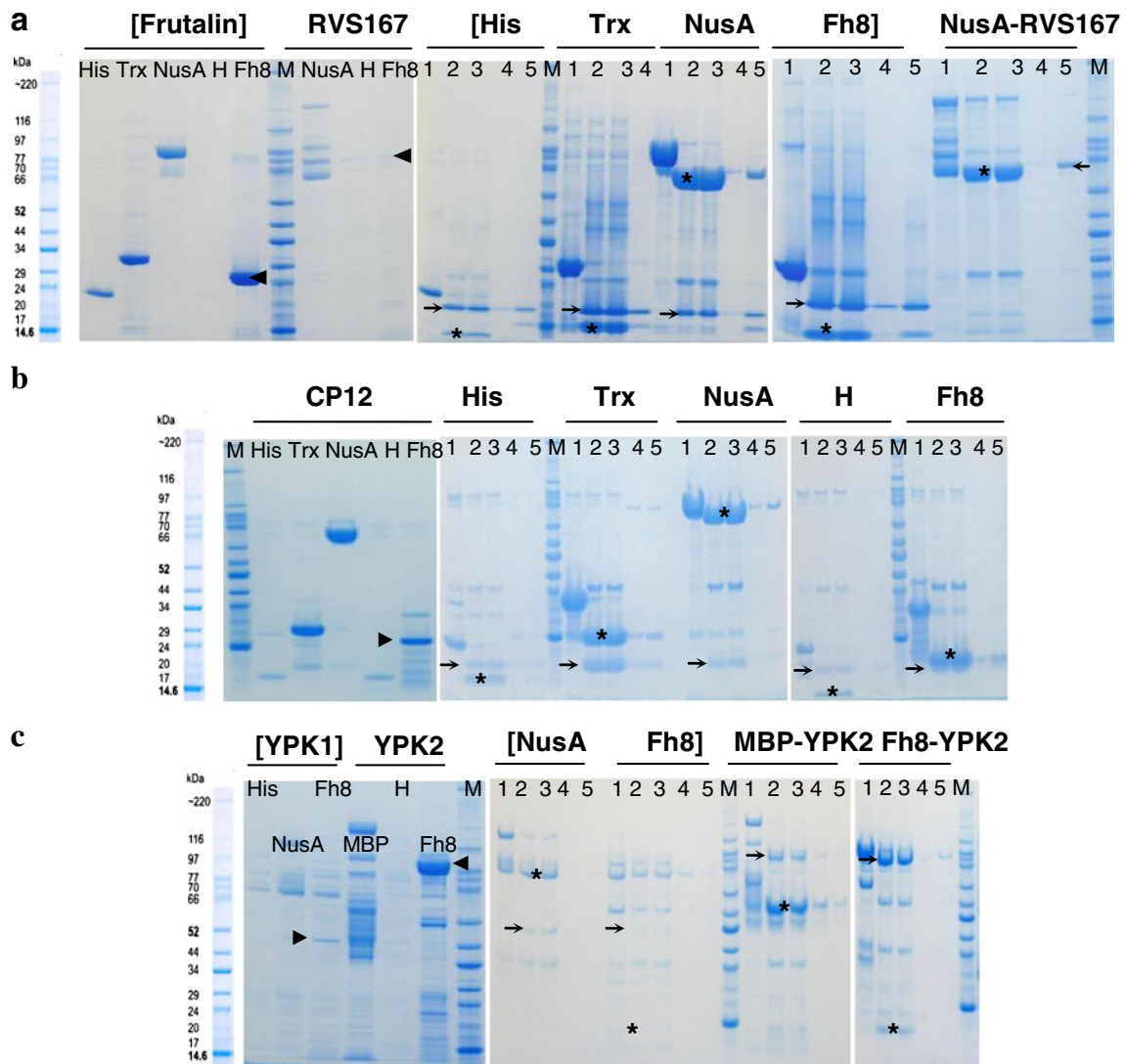
supernatant (soluble) sample. Arrows indicate the expected molecular weights for all fusion proteins following the fusion tags order loaded on the gels





**Fig. 3** Fusion proteins nickel-affinity purification: small-scale processing. **a** SDS-PAGE analysis of nickel affinity purifications. Gels were loaded following this sequence: supernatant fraction–flow through–washing step–elution 1–elution 2. Arrows indicate the location of fusion proteins at the observed/expected molecular weight.

**b** Soluble expression comparison per fusion tag of the different groups of target proteins. The “tag soluble expression” refers to the percentage of proteins with scores 1–3, estimated by Eq. 2. The “tag soluble score 2+3” refers to the percentage of proteins with scores 2 and 3, estimated by Eq. 3



**Fig. 4** Fusion proteins nickel-affinity purification and tag removal: semi-automated processing. **a** Frutalin (*in brackets*) and RVS167 5th eluted fractions and corresponding samples before/after TEV digestion and after cleaved protein purification. **b** CP12 5th eluted fractions and corresponding samples before/after TEV digestion and after cleaved protein purification. **c** YPK1 (*in brackets*) and YPK2 4th eluted fractions and corresponding samples before/after TEV digestion and after

cleaved protein purification. *M* In-house protein marker; 1–3 samples before TEV, after TEV, and after TEV plus centrifugation; 4 and 5 flow through and washing samples from the purification of the cleaved proteins. The *right arrowhead marks* show the Fh8-fusion proteins on the gel. *Arrows* indicate the expected molecular weight of cleaved proteins. *Asterisk* position the fusion tags after TEV digestion

than the NusA-containing cleaved protein. The RVS167 (Fig. 4a) and YPK1 (Fig. 4c) target proteins were obtained in higher soluble amounts when fused with the NusA tag than with the Fh8 tag. After tag removal, the cleaved YPK1 protein (35 kDa) from the NusA fusion yielded higher amounts than the one from the Fh8 fusion (Fig. 4c). For the RVS167 tag removal experiment, only the NusA fusion was tested. The SDS-PAGE bands resulting from this digestion (Fig. 4a) were difficult to distinguish as the molecular weight expected for the cleaved RVS167 (54 kDa) was close to the one (57 kDa) expected for the NusA. The YPK2 target protein, of which the soluble expression was not easy to assess in the small-scale screening, presented interesting

results in the scale-up experiments. Upon scale-up, this target protein revealed to be highly soluble expressed when fused to the Fh8 tag, out-performing the MBP fusion protein. The final cleaved YPK2 (78 kDa) from the Fh8 fusion was obtained in 2.3-fold higher amount than the cleaved protein from the MBP fusion.

## Discussion

In this work, a novel fusion system for soluble protein over-expression in *E. coli*—the Fh8 tag (8 kDa) and H tag (1 kDa)—is presented and compared to the traditional fusion

partners using a screening methodology. For the soluble protein expression screening, both Fh8 and H partners were inserted into the pETM vector series (Dummler et al. 2005). Using the same backbone, promoter, and cloning procedures, it was possible to achieve a systematic and consistent comparative analysis of the different solubility tags. All the selected pETM vectors are identical, differing only at the N-terminal fusion partners. Here, any differences found in target protein expression levels will probably be caused by the fusion partners sequence specific properties. The N-terminal position of the fusion partners seems to be a good option for optimal protein expression compared to the C-terminal position (not tested in this study) as it will allow the fusion partner to be translated first, providing time for the correct folding of the target proteins (Dyson et al. 2004). The Trx, GST, MBP, and NusA fusion partners were used in the screening comparison due to their known expression and solubility enhancing features and also because they are the most widely used fusion partners for recombinant protein production. The SUMO partner (Marblestone et al. 2006; Malakhov et al. 2004) was determinant for the selection of four target proteins, which were previously insoluble with this tag. Nevertheless, it was also included in this study to evaluate its solubility effect among the other fusion tags in the Frutalin and CP12 target proteins. The solubility of SUMO–Frutalin fusion was, however, not well evaluated probably due to a technical problem during the eight-tip sonication lysis, in which the ultrasound may not be equally distributed among the eight tips, resulting in the lower protein content observed in the small-scale screening of this fusion protein. The first expression and solubility comparisons between proteins fused to the Fh8 and H fusion partners and to the other tags were conducted in a small-scale screening, using qualitative SDS-PAGE and Bradford analyses. The small-scale screening strategy already proved to be a reliable tool for the comparison and selection of soluble proteins among different constructs, and it is a useful indicator of the expected protein production amounts upon scale-up expressions (Berrow et al. 2006; Dummler et al. 2005). Gel bands of fusion proteins were scored according to the 0–3 scale used by Berrow et al. (2006) and Bird (2011). In this study, scores 1–3 were given to an increased protein solubility level (for purified and supernatant samples) and were also used to estimate the total expression levels of fusion proteins. At a small-scale screening, scores 2 and 3 proved to be more consistent solubility predictors than the score 1, which may not represent a proper soluble expression in scale-up cultures (Bird 2011). Thus, three different evaluations were conducted here: the comparison of total expression (scores 1–3), the comparison of soluble expression (scores 1–3) and the comparison of soluble proteins with scores 2 and 3. The small-scale comparisons showed that the Fh8 fusion partner stands among the well-described best fusion partners, MBP, NusA, and Trx, for soluble protein expression (Kohl et

al. 2008; Nallamsetty and Waugh 2006). Interestingly, the Fh8 fusions presented lower total expression levels than these known tags. This difference in the expression levels may explain the better solubility results of some proteins when fused to the Fh8 tag. In spite of presenting higher total expression levels than the His tag fusions and identical to the Fh8 fusions, the H tag performed poorly in the protein solubility analysis, not working as solubility enhancing tag in *E. coli*. The GST fusion partner was previously evaluated as a relatively poor tag in several comparison studies (Bird 2011; Hammarstrom 2006; Ohana et al. 2009), which was confirmed also in our study. In addition to the fusion partners comparison, four different expression strains and two induction conditions were also tested in the small-scale screening. The best induction condition revealed to be correlated with the molecular weight of target proteins. Proteins with high molecular weights performed better at lower induction temperatures, in contrast to the CP12 protein, the smallest molecular weight target studied, which performed better with an induction temperature of 30 °C for 3 h. In fact, a lower induction temperature will slow down the protein expression rate, promoting a less stressful environment to the cell for protein production. A slow translation rate may improve the correct folding of higher molecular weight proteins and consequently their solubility (Berrow et al. 2006; Pacheco et al. 2012; Sorensen and Mortensen 2005; Terpe 2006). Most of the fusion proteins presented higher solubility when expressed in *E. coli* strains engineered with extra copies of rare codons [BL21 (DE3) Codon Plus-RIL and Rosetta (DE3)], thus confirming the importance of protein expression optimization via its host cell (Makino et al. 2011; Pacheco et al. 2012; Vernet et al. 2011). The small-scale screening was reproducible as shown by the comparison screenings before and after protein purification, which is essential to assess the real protein solubility among the different fusion tags (Dummler et al. 2005). In general, results from the scale-up analysis were consistent with the small-scale screening: The Fh8 fusion partner is among the best expression and solubility enhancing tags (as NusA and Trx). The YPK2 target (not successfully expressed in small-scale screening) was an exception, resulting in an improved solubility when fused to the Fh8 tag at the scale-up expression. Thus, when conducting a small-scale high throughput analysis, one must be aware of the balance between the loss of some target proteins for scale-up and the gain in increased number of parallel evaluations (Berrow et al. 2006). The YPK1 and RVS167 target proteins were, effectively, difficult to express as soluble proteins even using the NusA or Fh8 tags. In both screening methodologies, the position of the His<sub>6</sub> tag did not interfere with the expression and purification results since fusion constructs with the affinity tag in the middle of the fusion partner and the TEV cleavage site (NusA and Trx) performed as well as the fusion constructs with the affinity tag in the N-terminus (Trx, MBP

and Fh8). In both small-scale and scale-up experiments, fusion partners were removed using the TEV protease, leaving a glycine residue in the N-terminal of target proteins. The Fh8-derived target proteins presented similar final solubility as the Trx-derived ones and, curiously, performed better than the target proteins cleaved from NusA and MBP fusions. In fact, larger fusion partners are good solubilizers, but their large size can lead to an overoptimistic evaluation of protein solubility and yield (Dyson et al. 2004; Hammarstrom 2006; Kapust and Waugh 1999; Shih et al. 2002), as shown by the different solubility results before and upon tag removal. Looking among the solubility fusion partners used in this work, the Fh8 tag has the lowest contribution on the final size of the fusion protein (see Online Resource 6), which can explain the apparent lower solubility of Fh8 fusions in comparison to NusA or MBP fusions. In this context, the YPK2 target protein was again a particular case, in which the solubility effect of the Fh8 tag over the MBP tag was noticed in both “before” and “after” tag removal experiments. The SDS-PAGE and Bradford screening methodology used for the assessment of soluble expression before and after protein purification as well as upon tag removal is protein dependent, requiring a calibration for each protein (Hammarstrom 2006). Even so, a good agreement between the two methodologies was obtained, indicating that it can be used to predict and compare the protein soluble expression levels among different constructs.

The novel Fh8 fusion partner presented in this work revealed to be an effective tool for the improvement of protein solubility in *E. coli*. The conducted study pointed that (1) there is no “the best tag” for protein soluble expression, so multiple tags need to be tested with different proteins; (2) larger tags usually result in higher production yields, but these can lead to overestimation of the amount of soluble protein; (3) the smaller the size of the fusion tag, the easier it is to assess the solubility of the target protein. Thus, the Fh8 tag is an excellent candidate for testing expression and solubility next to the other well-known fusion tags. Its low molecular weight and its solubility enhancing effect make Fh8 an advantageous option compared to larger fusion tags for soluble protein production in *E. coli*.

**Acknowledgments** The financial support of the EMBL Heidelberg, Germany and Fundação para a Ciência e Tecnologia (FCT), Portugal, is acknowledged: the fellowship SFRH/BD/46482/2008 to Sofia J. Costa and the project PTDC/CVT/103081/2008. The authors wish to acknowledge Anne-Claude Gavin for providing four of the constructs for this study (RVS167, SPO14, YPK1, and YPK2) and Emmanuel Poilpré for the experimental help (both from the EMBL Heidelberg, Germany).

**Competing interests** The Fh8 tag utilization for the improvement of protein soluble expression in *E. coli* is covered by a worldwide patent (WO 2010082097) licensed to Hitag Biotechnology, Lda. The authors S.C., A.A., and A.C. are coowners of the patent and are associated with Hitag Biotechnology, Lda.

## References

- Berrow NS, Bussow K, Coutard B, Diprose J, Ekberg M, Folkers GE, Levy N, Lieu V, Owens RJ, Peleg Y, Pinaglia C, Quevillon-Cheruel S, Salim L, Scheich C, Vincentelli R, Busso D (2006) Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. *Acta Crystallogr D* 62:1218–1226. doi:10.1107/S0907444906031337
- Bird LE (2011) High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. *Methods* 55(1):29–37. doi:10.1016/j.ymeth.2011.08.002
- Davis GD, Elisee C, Newham DM, Harrison RG (1999) New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol Bioeng* 65(4):382–388. doi:10.1002/(SICI)1097-0290(19991120)65:4<382::AID-BIT2>3.0.CO;2-I
- De Marco V, Stier G, Blandin S, de Marco A (2004) The solubility and stability of recombinant proteins are increased by their fusion to NusA. *Biochem Bioph Res Co* 322(3):766–771. doi:10.1016/j.bbrc.2004.07.189
- Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* 27(3):297–306. doi:10.1016/j.biotechadv.2009.01.008
- Dummler A, Lawrence AM, de Marco A (2005) Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb Cell Fact* 4:34. doi:10.1186/1475-2859-4-34
- Dyson MR, Shadbolt SP, Vincent KJ, Perera RL, McCafferty J (2004) Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC Biotechnol* 4:32. doi:10.1186/1472-6750-4-32
- Esposito D, Chatterjee DK (2006) Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotech* 17(4):353–358. doi:10.1016/j.copbio.2006.06.003
- Hammarstrom M (2006) Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *J Struct Funct Genomics* 7:1–14. doi:10.1007/s10969-005-9003-7
- Kapust RB, Waugh DS (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* 8(8):1668–1674. doi:10.1110/ps.8.8.1668
- Kohl T, Schmidt C, Wiemann S, Poustka A, Korf U (2008) Automated production of recombinant human proteins as resource for proteome research. *Proteome Sci* 6:4. doi:10.1186/1477-5956-6-4
- LaVallie ER, Lu ZJ, Diblasio-Smith EA, Collins-Racie LA, McCoy JM (2000) Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. *Method Enzymol* 326:322–340. doi:10.1016/S0076-6879(00)26063-1
- Makino T, Skretas G, Georgiou G (2011) Strain engineering for improved expression of recombinant proteins in bacteria. *Microb Cell Fact* 10:32. doi:10.1186/1475-2859-10-32
- Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR (2006) Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci* 15(1):182–189. doi:10.1110/ps.051812706
- Malakhov MP, Mattern M, Malakhova OA, Drinker M, Weeks SD, Butt T (2004) SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics* 5:75–86. doi:10.1023/B:JSFG.0000029237.70316.52
- Nallamsetty S, Waugh DS (2006) Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expres Purif* 45(1):175–182. doi:10.1016/j.pep.2005.06.012
- Ohana RF, Encell LP, Zhao K, Simpson D, Slater MR, Urh M, Wood KV (2009) HaloTag7: a genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein



- purification. *Protein Express Purif* 68(1):110–120. doi:[10.1016/j.pep.2009.05.010](https://doi.org/10.1016/j.pep.2009.05.010)
- Oliveira C, Costa S, Teixeira JA, Domingues L (2009) cDNA cloning and functional expression of the alpha-D-galactose-binding lectin frutalin in *Escherichia coli*. *Mol Biotechnol* 43(3):212–220. doi:[10.1007/s12033-009-9191-7](https://doi.org/10.1007/s12033-009-9191-7)
- Oliveira C, Nicolau A, Teixeira JA, Domingues L (2011) Cytotoxic effects of native and recombinant frutalin, a plant galactose-binding lectin, on HeLa cervical cancer cells. *J Biomed Biotechnol* 2011. doi:[10.1155/2011/568932](https://doi.org/10.1155/2011/568932)
- Pacheco B, Crombet L, Loppnau P, Cossar D (2012) A screening strategy for heterologous protein expression in *Escherichia coli* with the highest return of investment. *Protein Express Purif* 81(1):33–41. doi:[10.1016/j.pep.2011.08.030](https://doi.org/10.1016/j.pep.2011.08.030)
- Peti W, Page R (2007) Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Express Purif* 51(1):1–10. doi:[10.1016/j.pep.2006.06.024](https://doi.org/10.1016/j.pep.2006.06.024)
- Sachdev D, Chirgwin JM (2000) Fusions to maltose-binding protein: control of folding and solubility in protein purification. *Method Enzymol* 326:312–321. doi:[10.1016/S0076-6879\(00\)26062-X](https://doi.org/10.1016/S0076-6879(00)26062-X)
- Shih YP, Kung WM, Chen JC, Yeh CH, Wang AHJ, Wang TF (2002) High-throughput screening of soluble recombinant proteins. *Protein Sci* 11(7):1714–1719. doi:[10.1110/ps.0205202](https://doi.org/10.1110/ps.0205202)
- Silva E, Castro A, Lopes A, Rodrigues A, Dias C, Conceicao A, Alonso J, da Costa JMC, Bastos M, Parra F, Moradas-Ferreira P, Silva M (2004) A recombinant antigen recognized by *Fasciola hepatica*-infected hosts. *J Parasitol* 90(4):746–751. <http://dx.doi.org/10.1645/GE-136R>
- Smith DB (2000) Generating fusions to glutathione S-transferase for protein studies. *Method Enzymol* 326:254–270. doi:[10.1016/S0076-6879\(00\)26059-X](https://doi.org/10.1016/S0076-6879(00)26059-X)
- Smith DB, Johnson KS (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* 67:31–40. doi:[10.1016/0378-1119\(88\)90005-4](https://doi.org/10.1016/0378-1119(88)90005-4)
- Sorensen HP, Mortensen KK (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb Cell Fact* 4:1. doi:[10.1186/1475-2859-4-1](https://doi.org/10.1186/1475-2859-4-1)
- Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Express Purif* 41(1):207–234. doi:[10.1016/j.pep.2005.01.016](https://doi.org/10.1016/j.pep.2005.01.016)
- Terpe K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biot* 60(5):523–533. doi:[10.1007/s00253-002-1158-6](https://doi.org/10.1007/s00253-002-1158-6)
- Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biot* 72(2):211–222. doi:[10.1007/s00253-006-0465-8](https://doi.org/10.1007/s00253-006-0465-8)
- Vernet E, Kotzsch A, Voldborg B, Sundstrom M (2011) Screening of genetic parameters for soluble protein expression in *Escherichia coli*. *Protein Express Purif* 77(1):104–111. doi:[10.1016/j.pep.2010.11.016](https://doi.org/10.1016/j.pep.2010.11.016)
- Waugh DS (2005) Making the most of affinity tags. *Trends Biotechnol* 23(6):316–320. doi:[doi:10.1016/j.tibtech.2005.03.012](https://doi.org/10.1016/j.tibtech.2005.03.012)
- Yao L, Yin J, Zhang X, Liu Q, Li J, Chen L, Zhao Y, Gong P, Liu C (2007) *Cryptosporidium parvum*: identification of a new surface adhesion protein on sporozoite and oocyst by screening of a phage-display cDNA library. *Exp Parasitol* 115(4):333–8. doi:[10.1016/j.exppara.2006.09.018](https://doi.org/10.1016/j.exppara.2006.09.018)