Optimized expression of the Starmerella bombicola lactone esterase in Pichia pastoris through temperature adaptation, codon-optimization and co-expression with HAC1

De Waele S.1; Vandenberghe I.1, Laukens B.1,2, Planckaert, S.1, Verweire, S.3, Van Bogaert I.N.A.4, Soetaert W.3, Devreese B.1,*, Ciesielska K.1

1Laboratory for Protein Biochemistry and Biomolecular Engineering, Ghent University, Ghent, Belgium
2Unit for Medical Biotechnology, Center for Medical Biotechnology (CMB), VIB, Ghent, Belgium
3Centre for Synthetic Biology, Ghent University, Ghent, Belgium
4Laboratory of Industrial Biotechnology and Biocatalysis, Ghent University, Ghent, Belgium

*Corresponding author

Bart Devreese
Laboratory for Protein Biochemistry and Biomolecular Engineering
Department of Biochemistry and Microbiology
Ghent University

KL Ledeganckstraat 35
B-9000 Gent
Belgium

Bart.Devreese@ugent.be
Abstract

The *Starmerella bombicola* lactone esterase (SBLE) is a novel enzyme that, *in vivo*, catalyzes the intramolecular esterification (lactonization) of acidic sophorolipids in an aqueous environment. In fact, this is an unusual reaction given the unfavorable conditions for dehydration. This characteristic strongly contributes to the potential of SBLE to become a 'green' tool in industrial applications. Indeed, lactonization occurs normally in organic solvents, an application for which microbial lipases are increasingly used as biocatalysts. Previously, we described the production of recombinant SBLE (rSBLE) in *Pichia pastoris* (syn. *Komagataella phaffii*). However, expression was not optimal to delve deeper into the enzyme’s potential for industrial application. In the current study, we explored codon-optimization of the *SBLE* gene and we optimized the rSBLE expression protocol. Temperature reduction had the biggest impact followed by codon-optimization and co-expression of the *HAC1* transcription factor. Combining these approaches, we achieved a 32-fold improvement of the yield during rSBLE production (from 0.75 mg/L to 24 mg/L culture) accompanied with a strong reduction of contaminants after affinity purification.

Keywords

Lipase; lactonase; *Starmerella bombicola*; *Pichia pastoris*; protein purification; Green chemistry.
Highlights

- Production of *Starmerella bombicola* lactone esterase (SLE) in *P. pastoris* is evaluated.
- Expressing rSLE at 16°C decreased the amount of contaminants during purification.
- Codon-optimization and expressing at 16°C increased rSLE yield seventeenfold.
- Co-expression of rSLE with *HAC1* increased the yield approximately twofold.

Abbreviations

rSLE: Recombinant *Starmerella bombicola* lactone esterase; rSLEopt: Recombinant *Starmerella bombicola* lactone esterase obtained after codon-optimization; EndoH: Endoglycosidase H; SEC: Size-exclusion chromatography; IMAC: immobilized metal-ion affinity chromatography; MALDI-TOF: Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight; UPR: unfolded protein response
Introduction

Microbial lipases are widely used as biocatalysts in a variety of biotechnological applications[1]. Although the typical natural activity of lipases in aqueous media is the hydrolysis of triglycerides or phospholipids, these enzymes can also catalyze inter- and intra-esterifications in organic media. This is used for the synthesis of intermediates for the production of chiral compounds such as many pharmaceuticals[1]. Intra-esterifications result in the formation of lactones, which occur widely in nature as hormones (spironolactones and mevalonolactones), antibiotics (erythromycin and amphotericin B) or neurotransmitters (butyrolactones and avermectins). Intermolecular esterification occurs also in the biosynthesis of sophorolipids (SLs), a family of fungal biosurfactants that naturally exist in two forms: a closed lactone and an open acidic form[2]. Each form has distinct properties: acidic SLs have better foaming properties, while lactonic SLs are better in surface tension reduction and have antimicrobial activity[3].

We have previously described that ring closure to form lactonic SLs in *S. bombicola* is catalyzed by a novel member of the *Pseudozyma (Candida) antarctica* – A (CAL-A) lipase family[4], which was designated as the *Starmerella bombicola* lactone esterase (SBLE)[5]. This enzyme is rather unique as it lactonizes sophorolipids in the overabundance of water[5, 6], whereas normally lipases perform esterifications in non-aqueous media[7]. Indeed, only a few enzymes are described to be capable to perform esterification reactions in an aqueous environment[8, 9]. At present, lactonization reactions are mostly performed using a different lipase of *Pseudozyma (Candida) antarctica*, the *Pseudozyma (Candida) antarctica* lipase B (CAL-B), commercialized as Novozyme® 435 by Novozymes. CAL-B has been used for the synthetic preparation of lactonized SLs, but hazardous solvents (dry tetrahydrofuran (THF) containing vinyl acrylate and vinyl acetate) are required [10]. In addition, CAL-B links the fatty acid moiety to the C-6” glycosyl hydroxyl group whereas the natural reaction, as performed by SBLE, involves the C-4” hydroxyl group[10].

The unique properties of SBLE could provide an alternative for the catalysis of lactonization reactions in green chemistry applications[6, 11]. However, in order to valorize SBLE for industrial use, a better production system is required. We have previously reported a *Pichia pastoris* production system for the production of rSBLE (Figure 1, Table 1).
Figure 1: Comparison of rSBLE (left) and SBLE (right). We replaced the endogenous secretion signal used in S. bombicola with Saccharomyces cerevisiae’s α-factor secretion signal. This leaves two additional amino acids (AG) at the N-terminus of the protein. At the C-terminal we added a His6 tag. In addition two amino acids are inserted (underlined) to allow restriction site compatibility. The structure displayed is a model based on the structure of CAL-A (PDB id: 2VEO) [12].

Table 1: Characteristics of the Starmerella bombicola lactone esterase.

<table>
<thead>
<tr>
<th></th>
<th>rSBLE (SBE-His)</th>
<th>SBLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular weight (kDa) *</td>
<td>45.122</td>
<td>44.085</td>
</tr>
<tr>
<td>Subunits/isoforms</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>pI (theoretical)</td>
<td>4.88</td>
<td>4.63</td>
</tr>
<tr>
<td>pH optimum</td>
<td>3.5 - 6.0</td>
<td>Not determined</td>
</tr>
<tr>
<td>Temperature optimum</td>
<td>40°C</td>
<td>Not determined</td>
</tr>
</tbody>
</table>

Table 1: Characteristics of the Starmerella bombicola lactone esterase.

Temperature and pH optima have been retrieved from [6], while pI and molecular weight have been obtained from ExPASY [13]. *: the influence of glycosylation on the molecular weight has been excluded due to the uncertainty of the actual amount of N-glycosylation sites.

Unfortunately, this system was so far unsatisfactory for two main reasons: first, only a low yield of rSBLE (0.75 mg/liter culture) was obtained. Second, during His Tag purification, several contaminants were co-purified. Therefore this study foremost aimed to improve the rSBLE production system in view of further exploitation of this enzyme. To improve the production of rSBLE we followed a dual approach. On the one hand, we optimized the expression conditions using the strain expressing the non-codon-optimized SBLE (containing the plasmid pPICZαB_rSBLE)[6] and on the other hand, we generated a novel, codon-optimized rSBLE production strain. In brief, we investigated the
influence of aeration, pH, induction-time, methanol feeding and temperature on rSBLE accumulation in the culture supernatant. Afterwards, these optimized expression conditions were implemented on the codon-optimized construct to reach large-scale expressions. Additionally, we verified whether co-expression of the transcriptional activator gene HAC1 improved production yields. Indeed, during expression of recombinant protein, proteins can become unfolded leading to the unfolded protein response (UPR). The transcriptional activator Hac1p is produced during this unfolded protein response, stimulating the transcription of several genes related to translocation, glycosylation and protein folding[14, 15]. Co-expression of Hac1p has been used successfully to increase the yield of the protein of interest, due to the recovery of unfolded protein[15]. Overall, our efforts led to a significant improvement of production yields for SBLE production.

**Materials and methods**

**Strains and media:**

The strains used in this study are shown in Table 2. *Escherichia coli* DH5α (New-England-Biolabs (NEB)) was used for cloning and plasmid amplification. Bacteria were propagated in low-salt lysogeny broth (LS-LB) medium, consisting of 0.5% (w/v) sodium chloride (Merck), 0.5% (w/v) yeast extract (Lab M) and 1.0% (w/v) tryptone (Lab M) with or without 1.5% agar (BD) and with the required antibiotics. The *Pichia pastoris* (syn. *Komagataella phaffii*) NRRL-Y-11430 strain (kindly obtained from Prof. Nico Callewaert (VIB, Ghent University)) was used for recombinant protein production. Yeast strains were plated on yeast extract peptone dextrose (YPD) plates (1% (w/v) yeast extract, 2% (w/v) peptone (BD), 2% (w/v) dextrose (Merck), 1.5% agar (Difco)) with the required antibiotics. All antibiotics were purchased from Thermo Fisher Scientific with the exception of carbenicillin, which was obtained from Gold Biotechnology. For recombinant protein expression, strains were grown in buffered glycerol-complex medium (BMGY) and induction was performed in buffered methanol-complex medium (BMMY). Both BMGY and BMMY consist of 1% (w/v) yeast extract, 2% (w/v) peptone (BD), 100 mM phosphate buffer (Sigma-Aldrich) at pH 6.0 and 1.34% (w/v) yeast nitrogen base (YNB, Formedium) with 1% glycerol (v/v Sigma) or 1% methanol (v/v, VWR) as sole carbon source respectively.
**Cloning of the codon optimized rSBLE**

All enzymes required for cloning were obtained from NEB. The original DNA construct for rSBLE has been described in [6]. In brief, the coding sequence of mature SBLE (GenBank accession no. **JB750219**), lacking the secretion signal and the stop codon, was cloned in-frame with *Saccharomyces cerevisiae*’s prepro α-mating factor secretion signal and a C-terminal His-tag. To determine the impact of possible codon bias in *P. pastoris*, the SBLE ORF was codon-optimized by Genscript®’s proprietary algorithm and ordered synthetically. The codon-optimized SBLE construct was cloned to the pPICZαB plasmid (Invitrogen®, Carlsbad, USA) (Table 2; sequence in supplementary Figure 1). Therefore, the SBLE construct was digested with PstI-HF® and SalI-HF® and ligated to the similarly opened pPICZαB plasmid using T4 DNA ligase. The ligation reaction mixture was used to transform electrocompetent DH5α cells and positive transformants were selected on LS-LB plates containing Zeocin™ (50 µg/ml medium). Twelve positive clones were selected and screened by colony PCR using primers FW_rSBLE and Rev_rSBLE (5’-ATTGCTGACGGACTCCCTTGGGTTAT-3’ and 5’-ATTAGTCGACTGTGGGCTAGAATTAACTGG-3’, respectively. The underlined sequence indicates the PstI- and SalI-sites). A single clone was selected, the resulting plasmid (pPICZαB_rSBLEopt) was isolated and sequence-verified by GATC biotech (Konstanz, Germany) using the primers mentioned above.

### Table 2: Plasmids and strains used in this study.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Characteristics</th>
<th>Size (bp)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pPICZαB</td>
<td>Zeo®, α-factor secretion signal, His6 Tag, C-myc Epitope tag, AOX1 promoter, 3’ AOX TT</td>
<td>3597</td>
<td>Life Technologies®</td>
</tr>
<tr>
<td>pPICZαB_rSBLE</td>
<td>Zeo®, α-factor secretion signal, His6 Tag, AOX1 promoter, 3’ AOX TT, linearized with SacI, coding original rSBLE</td>
<td>4678</td>
<td>[5]</td>
</tr>
<tr>
<td>pPICZαB_rSBLEopt</td>
<td>Zeo®, α-factor secretion signal, His6 Tag, AOX1 promoter, 3’ AOX TT, linearized with SacI, coding codon-optimized rSBLE</td>
<td>4678</td>
<td>This study</td>
</tr>
</tbody>
</table>
Transformation of codon-optimized rSBLE and co-expression vectors

The plasmid of interest (pPICZαB_rSBLEopt) was linearized in the AOX1 promoter region by SacI and 100 ng of the plasmid was transformed to *P. pastoris* NRRL-Y-11430 cells using electroporation, as in the protocol of Wu et al. [17]. Transformants were selected on YPD medium containing Zeocin™ (500 µg/ml) and the best producing clone was chosen based on Western blot analysis.

To study the influence of Hac1p co-expression, the best producing rSBLE clone was made to be electrocompetent using the protocol of Wu et al. [17]. Subsequently, 100 ng of pPIC9K_HAC1spliced, linearized by PmeI in the AOX1 promoter region, was added to these electrocompetent cells. Transformants were then selected on YPD medium supplemented with Hygromycin B (300 µg/ml) and Zeocin™ (500 µg/ml).

Evaluation of the effect of aeration using baffled flasks

Cells, containing the plasmid of interest, were grown in a preculture of 10 ml BMGY overnight at 250 rpm, while maintaining Zeocin™ (500 µg/ml) selection. This preculture was divided between two types of shake flasks, both containing 500 ml BMGY, and grown for 48 hours to evaluate the effect of aeration. The first type

<table>
<thead>
<tr>
<th>Strains</th>
<th>Characteristics</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRRL-Y-11430</td>
<td><em>Pichia pastoris</em> (syn. <em>Komagataella phaffii</em>) strain, used for expressing protein of interest</td>
<td>Prof. Nico Callewaert (VIB, Ghent University)</td>
</tr>
<tr>
<td>DH5α</td>
<td>F-, Φ80lacZΔM15 ΔlacZYA-argF) U169 recA1 endA1 hsdR17 (rK-, mK+) phoA supE44 λ- thi-1 gyrA96 relA1, used for producing the plasmids of interest and cloning purposes</td>
<td>NEB®</td>
</tr>
</tbody>
</table>

Amp, ampicillin; AOX, alcohol oxidase; bp, basepairs; His, histidine; Hyg, hygromycin; opt, optimized; R, resistant to; rSBLE, recombinant 135*Starmerella bombicola* lactone esterase; TT, transcription terminator; Zeo, Zeocin™.
0735-10-3000) did not have any baffles, while the latter shake flask did contain 4 baffles (Neubert-Glas, 0749-10-3000). Both flasks types had a volume of 3000 ml. After 48 hours of induction the cells were centrifuged and the supernatant was studied regarding rSBLE production.

Screening for optimal conditions for expression of rSBLE.

Cells were grown in a 24-well plate sealed with AreaSeal film (Sigma-Aldrich). Initially, each well contained 2 ml BMGY. Cells were allowed to grow for 48 hours (28°C, 250 rpm) to accumulate biomass. Cells were centrifuged, and washed. Afterwards, the cells resuspended in BMMY for induction under different expression conditions to screen for the effect of pH, methanol-feeding and induction-time. The following parameters were tested: pH (0.1 M citrate buffer, pH 2.0, 3.0, 4.0, 5.0, 6.0; 0.1 M phosphate buffer, pH 6.0), percentage of methanol for induction (0.5%, 0.75%, 1.0% and 1.5%), induction-time (48h, 60h, 72h, 84h, 96h, 102h and 120h) and induction-temperature (16°C and 28°C). The results were always compared with the reference condition as used in [6] where expression conditions were: pH 6.0 (0.1 M phosphate buffer), temperature 28°C, duration 48h and methanol concentration 1%. After the appropriate induction time, the cultures were centrifuged (6000 g) and the supernatant was collected for analysis. The cleared supernatant was precipitated using DOC-TCA precipitation. In brief, 0.05% (w/v, final concentration) sodium deoxycholate (DOC, Sigma-Aldrich) was added to 1 ml of supernatant. After incubation, 20% (w/v, final concentration) of trichloroacetic acid (TCA, Sigma-Aldrich) was added and the sample was incubated on ice. Subsequently, the sample was centrifuged and the obtained pellet was washed twice with ice-cold acetone. After washing with 70% (v/v) ethanol, the pellet was resuspended in 50 µl phosphate buffered saline (PBS). Prior to SDS-PAGE analysis, 5 µl of sample was treated with Endoglycosidase H (Endo H, 500 U, NEB) in order to remove heterogeneity due to N-glycosylation and to allow a more accurate molecular weight estimation on SDS-PAGE.

Protein analysis: SDS-PAGE and Western blotting.
Protein samples were analyzed by SDS-PAGE on 12.5% gels (Tris-HCl). The Precision Plus Protein™ Unstained Standard (Bio-Rad) was used as a molecular weight marker. Gels were stained with Coomassie brilliant blue G, unless Western blotting was used.

For Western blot analysis of rSBLE, rSBLE was detected using an Anti-His Tag horseradish peroxidase-coupled antibody (Anti-His (C-Term)-HRP, 46-0707, Invitrogen, Carlsbad, USA). Blots were revealed using luminol and H₂O₂ (Pierce ECL Western blotting substrate, both obtained from Thermo Scientific) and using the Precision Plus Protein™ Dual Color Standard (Bio-Rad). The gels were scanned using a GS-800 calibrated densitometer (Bio-Rad) and visualized using the Quantity One software package (Bio-Rad).

**Protein purification**

For purification of rSBLE, a pre-culture was prepared in 10 ml BMGY and grown overnight (at 28°C, 250 rpm) under antibiotic selection. The next day, the pre-culture was used to inoculate 500 ml BMGY in 3L baffled flasks (at 28°C, 250 rpm). At an optical density at 600 nm (OD₆₀₀) of approximately 41, cells were harvested by centrifugation (10 minutes at 4,000 g). The cells were resuspended with BMMY containing 1% methanol (v/v) to induce expression. To maintain the induction, 1% methanol (final concentration) was added to the cells every 10 to 12 hours. After 48 hours of induction, the cultures were centrifuged and the supernatant was harvested.

All purifications were done on an ÄKTA Purifier system (GE Healthcare). Prior to loading the sample, 100 mg/l reduced glutathione (Sigma-Aldrich) and 2 mM (final concentration) of magnesium sulfate (Sigma-Aldrich) were added and the pH was adjusted to 7.5. Any precipitation was removed by filtering the sample over a 0.22 µm bottle top filter (Sarstedt). The cleared supernatant was loaded overnight on a 5 ml HisTrap™ FF column (GE Healthcare) previously equilibrated with washing buffer (50 mM Na₂HPO₄ and 500 mM NaCl (pH 7.5)). The column was washed with washing buffer until the UV(280nm) absorption reached baseline. The column was step-wise eluted with 20-, 200 and 400 mM imidazole in washing buffer. rSBLE eluted at 20 mM imidazole already, but a significant amount was recovered at 200mM as well. These two fractions were concentrated by ultrafiltration to a volume of 1.0 ml using 10
kDa molecular weight cut-off Vivaspin® columns (EMD Millipore). Ten µl of these concentrated fractions was analyzed by SDS-PAGE as described above.

To further purify rSBLE, the pooled and concentrated IMAC fraction was injected onto a HiLoad® 16/600 Superdex® 200pg column (GE Healthcare) equilibrated with 150 mM NaCl in 50 mM Tris (pH 7.5) and eluted with the same buffer. The fractions containing rSBLE were concentrated to 1.0 ml by ultrafiltration using Vivaspin® (10 kDa molecular weight cut-off, Merck millipore) from which 10 µl was taken for SDS-PAGE analysis. The size-exclusion chromatogram was represented as a graph made by GraphPad prism 6.0. Every purification was performed with at least 2 technical replicates and the average ± standard deviation of the total yield was used to compare with other conditions. The obtained concentrated SEC fraction was stored at -80°C until further analysis.

**Determination of protein concentration**

The rSBLE concentration was determined by UV spectrometry with a NanoDrop® 2000 (Thermo Scientific) using parameters: ε, 73.355 M⁻¹ cm⁻¹ and a molecular weight of 45.5 kDa as calculated using ProtParam (ExPASY[13]).

**Activity assays of rSBLE**

The activity of rSBLE, isolated from different productions was analyzed using an HPLC-based activity assay following the protocol described in Ciesielska et al [6]. In brief, 2 µg of purified rSBLE was added to 500 µl of reaction buffer, containing 5 mM of acidic diacetylated sophorolipids and 50 mM sodium citrate. The mixture was incubated for 1 hour at 1400 rpm after which reaction was stopped using 500 µl 100% (v/v) EtOH. After concentrating the sample using a SpeedVac vacuum centrifuge (Thermo Savant, Holbrook, NY) to 120 µl, the samples were analyzed using HPLC.

**Results**

When rSBLE is produced in *P. pastoris* as described previously [6] it displays a heterogeneous pattern on SDS-PAGE analysis (Figure 2A) and on Western blotting (Figure 2B, right lane). In order to more accurately estimate rSBLE quantity, samples were deglycosylated prior to SDS-PAGE analysis. rSBLE appears, after deglycosylation, as a
predominant band with an apparent molecular weight of approximately 50 kDa and a minor band was observed around 37 kDa (Figure 2AB).

Influence of aeration, pH, percentage of methanol, temperature and expression duration on rSBLE production.

First, we evaluated whether aeration of the cultures has an impact on rSBLE expression. To this end we compared the production yields obtained using Erlenmeyers with and without baffles. From these experiments we found that using baffled flasks (Neubert-Glas, 0749-10-3000) led to a modest increase in rSBLE expression (Figure 2B). Next, we tested whether altering the pH of the medium during induction would influence rSBLE expression compared to a standard phosphate-buffered medium at pH 6.0. When inducing the cultures in a citrate buffer pH 3.0, we observed a somewhat increased accumulation but this coincided with an increased degradation (Supplementary Figure 2). No effect on rSBLE accumulation was observed when the induction was performed pH of 2.0, 4.0 and 5.0 (results not shown). Based on these results, the standard phosphate buffer (pH 6.0) was maintained.

We also determined whether varying the methanol feed during induction could improve rSBLE induction compared to the use of 1.0% methanol. However, altering methanol feed did not have any impact rSBLE yields (results not shown). To determine the optimal induction time and to decipher whether rSBLE continues to accumulate in the culture medium, we compared various extended induction times (60h, 72h, 84h, 96h, 108h and 120h). The gel patterns of proteins obtained after DOC-TCA precipitation of the medium were compared to a previously obtained sample collected after 48h induction time. Although over time, there is a small increase in intensity of the band corresponding to rSBLE, we also observed this was also the case for several other proteins in the sample. Therefore, we concluded that increasing induction time would probably also compromise protein purification. (Figure 2C). Consequently, the 48 hours induction time was maintained in future experiments.

Finally, we determined the effect of the temperature used during expression. To this end, we first grew the cells at 28°C but then lowered the temperature to 16°C for induction. We also maintained a reference culture at 28°C during induction. After SDS-PAGE analysis, we observed that performing the induction at 16°C resulted in a considerable
increase of rSBLE, compared to induction at 28°C (Figure 2D). Curiously, based on the band intensity it seems that
the rSBLE yield at the 9 hour timepoint is almost as good as after 48 hours of induction. Nevertheless, Western blotting
does display a difference (Figure 2B). This conflicting result could be explained by assuming a limited sensitivity/dynamic range of Coomassie Brilliant Blue staining compared to immunoblotting.

Based on these experiments, we further performed protein productions in baffled shake flasks and the induction was
performed by feeding cultures with 1% methanol at pH 6.0 using a standard phosphate buffer at 16°C for 48 hours.

Figure 2: Optimization of rSBLE production. A) Coomassie Brilliant Blue stained SDS-PAGE of purified wild type rSBLE (WT) and deglycosylated rSBLE (DG). The star indicates Endo H® B) Western blot analysis showing the influence of aeration on rSBLE-accumulation. rSBLE production in non-baffled flask was compared to a production in baffled flask at different time points, indicated in hours after induction. Samples were deglycosylated prior to analysis. ‘+’ represents non deglycosylated rSBLE. C) Coomassie Brilliant Blue stained SDS-PAGE analysis to evaluate rSBLE production at prolonged induction times (60-120 hours). The reference (ST) is rSBLE obtained after 48 h induction time, a sample obtained from an independent experiment and its untreated, non-deglycosylated control (UN). D) Coomassie Brilliant Blue stained SDS-PAGE analysis to compare rSBLE production at different temperatures (16°C and 28°C). Cells were harvested after 48 hours post-induction. ‘-’ represents a control sample where rSBLE production was not induced by methanol addition. In all pictures, ‘M’ stands for marker, molecular weight is shown in kDa. The black arrow indicates deglycosylated rSBLE.
Influence of expression temperature on rSBLE yield and retained contaminants during purification

We reported that during protein purification of rSBLE, two dominant contaminants co-eluted in during IMAC purification (phosphotyrosine phosphatase and phosphatidylinositol-4-kinase) [6]. These proteins were presented in a low molecular weight smear on an SDS-PAGE gel, together with degradation products of rSBLE. We investigated whether these contaminants were present after purification of samples evolved from our improved induction protocol (Figure 3). When comparing the SDS-PAGE patterns of the rSBLE containing fractions after IMAC purification (Figure 3A), we noticed that the samples obtained from cultures induced at 16°C contained much less low molecular weight contaminants or rSBLE degradation (the ‘smearing’ around 25 kDa) than the cultures induced at 28°C. The dramatic reduction of protein degradation or copurification of contaminants is also observed when comparing the size-exclusion chromatograms of these IMAC elution fractions (Figure 3B). Not only did we observe less contaminants, we also observed an approximate sixfold increase of rSBLE yield after purification, from 0.75 mg/L rSBLE to 4.7 mg/L as determined by UV spectroscopy.

Figure 3: Comparison of rSBLE production expressed by the non-codon-optimized construct at 28°C and 16°C. A) Coomassie® Brilliant Blue stained SDS-PAGE patterns comparing the concentrated protein fractions (S) eluted from IMAC after imidazole elution. B) Size-exclusion chromatograms after loading the IMAC fraction from the original construct expressed at 16°C (solid line) and at 28°C (dashed line). The void volume, until 40 ml elution volume, is not shown. ‘Contaminants’ represent the co-eluting low molecular weight contaminants and possible rSBLE degradation products. M, molecular weight marker in kDa; mAU, milli absorbance units.

Evaluation of the codon-optimized rSBLE construct.
In parallel with the evaluation of the expression conditions described above, we produced a strain expressing a codon-optimized rSBLE construct. We isolated the rSBLE produced by this strain (rSBLEopt) using the same purification scheme, and compared the results with the product obtained using the previously described construct (Figure 4). We achieved a 3-fold increase in yield for the rSBLEopt sample, as yields went from 4.7±0.5 mg/l to 13.7±1.5 mg/l (Figure 4A). In addition, size-exclusion chromatography displayed a single peak with the codon-optimized constructs. The chromatogram of the non-codon-optimized construct showed a more heterogeneous pattern that we attributed to difference in glycosylation. (Figure 4B, [6]). We tested whether this led to a difference in activity using a HPLC assay, but this was not the case (Supplementary Figure 3).

Figure 4: Comparison of the expression of the original rSBLE construct (dashed line) and the codon-optimized rSBLE construct (solid line). A) Coomassie Brilliant Blue stained SDS-PAGE analysis of the fractions eluted from IMAC using imidazole (S, concentrated sample eluted from IMAC with imidazole). B) Size-exclusion chromatography analysis of both concentrated IMAC fractions. The void volume, with the exception of 40 to 45 ml, was not shown. The dashed line chromatogram represents the SEC profile of the expressed rSBLE from the original construct while the solid line chromatogram shows expressed rSBLE from the codon-optimized construct. M, molecular weight marker in kDa; mAU, milli absorbance units; S, concentrated IMAC sample.

The influence of co-expression of HAC1 on codon-optimized rSBLE's yield.

We finally tested the effect of co-expression with the transcriptional activator HAC1 on rSBLEopt yield. Activity of HAC1 in this strain was confirmed by verifying increase of production of Kar2p and Pdip, two proteins known to be affected by HAC1 activity (Supplementary file and Supplementary Figure 4) Effectively, this led to a dramatic improvement, i.e. the yield raised from 13.7±1.5 mg/l to 24.1±1.0 mg/l (Figure 5A). On the size-exclusion
chromatogram, the HAC1 co-expressed rSBLEopt eluted again as a single peak, at the same retention volume (Figure 5B). rSBLE, whether it was co-expressed with HAC1 or not, was tested for activity as was described in [6] but no difference was observed (Supplementary Figure 3).

Figure 5: Comparison of the expression of the codon-optimized rSBLE (dashed line) and the codon-optimized rSBLE (solid line) co-expressing HAC1. A) Coomassie Brilliant Blue stained SDS-PAGE analysis of both concentrated IMAC fractions (S, concentrated IMAC sample). B) The void volume, with the exception of 40 to 45 ml, was not shown. The dashed line chromatogram was the SEC profile of the expressed rSBLE from the codon-optimized construct while the solid line chromatogram represents codon-optimized rSBLE co-expressed with HAC1. M, molecular weight marker in kDa; mAU, milli absorbance units; S, concentrated IMAC sample.

While comparing the different obtained yields, one could argue that a lower temperature and codon-optimization led to a protein that can be more easily purified, and that improvement of yield is not necessarily caused by an increase of expression. To test this possibility, the 3 different strains producing rSBLE, rSBLEopt and rSBLEopt co-expressed with Hac1, were compared. After 48 hours induction, the obtained secreted proteins were collected using DOC-TCA precipitation. Protein yield was verified by Western Blotting. The results of this analysis proved that improvement of yield after purification is effectively the result of an increased protein production in the medium (Figure 6).
Figure 6: Western blot analysis of different production strains. Samples (1 ml of culture each) were precipitated using DOC-TCA precipitation. Bradford analysis was used to normalize the protein concentrations after which an equal amount of total protein was loaded. The black arrow represents deglycosylated rSBLE.

Discussion

In this study we aimed to optimize the expression conditions for rSBLE production in *P. pastoris* [5, 6]. It has been previously reported that the flask design is a crucial parameter in the fermentation design[18]. Therefore we first studied the effect of using a baffled shake flask on the expression of rSBLE. Western blotting showed an increase in band intensity of rSBLE. Indeed, baffled shake flasks typically result in a higher amount of cells and improved aeration, which can in turn lead to a higher protein production. It should be mentioned that the actual design of the baffled flasks can also have an effect on the protein production[18], but we only tested the four-baffled shake flask.

Another approach to increase the expression of target proteins is to optimize the expression conditions. More precisely, optimization of the cultivation temperature, methanol concentration, pH and the expression time can greatly increase protein production[19]. In the case of rSBLE, neither the increase of the expression time nor a variation of the methanol concentration influenced rSBLE expression, in contrast to different examples described in the literature[20]. While
expressing rSBLE at a pH of 3 led to an increase of band intensity, this coincided with an increase in degradation. This could be explained by a decrease in stability of the protein when exposed to low pH. However, during sophorolipid fermentation, pH typically drops to 3.5 which did not hinder SBLE’s activity [6]. Therefore, it seems more likely that the observed degradation is caused by proteolytic activity. *Pichia pastoris* is known to produce intracellular proteases and unwanted proteolysis can be largely solved by protein production in *pep4* (SMD1168) mutant cells, devoid of the major intracellular protease [21]. As a last parameter, we performed induction at 16°C instead of 28°C. By reducing the induction temperature, we decrease the rate of protein synthesis, and give time for newly translated recombinant protein to fold properly. According to other literature examples, lowering induction temperature does not only lead to a higher expression, it can also lead to a more active protein [22] and an increase in cell viability [23, 24]. Indeed, transcriptomic [25] and proteomic [26] analyses of *Pichia pastoris* recombinant protein expression at a lower temperature showed a strong decrease in folding stress, noticed by the decrease of chaperones and other folding-related proteins. Changing the expression temperature had the largest impact on production yield, raising it to 4.7±0.5 mg/l culture.

Besides its low yield, the rSBLE production as described so far was also compromised by an overabundance of contaminants which exhibit high affinity for the IMAC column. These contaminants were identified as intracellular proteins with an apparent molecular weight of 15 and 30 kDa according to SDS-PAGE, and were identified as a phosphotyrosine phosphatase (UniProt code: C4QXK8) and a fragment of phosphatidylinositol-4-kinase (UniProt code: C4QV87). Additionally, tryptic peptides from rSBLE were also identified from SDS-PAGE bands cut at lower molecular weight, indicating rSBLE degradation [6]. Interestingly, the identified contaminants were not yet described in the literature [27]. We observed that expressing rSBLE at 16°C not only increases the protein yield but also reduces the amount of such contaminants avoiding the need for additional clean-up steps.

Codon bias is another major barrier to obtain high protein yields. This codon bias occurs when foreign coding DNA is significantly different from that of the host. As a consequence, during the synthesis of the recombinant protein, depletion of low-abundance tRNAs can arise. This can lead to amino acid disincorporation or truncation of the polypeptide, thus affecting the heterologous protein production levels [28]. When we compared the codon usage of *P.
pastoris with the SBLE coding sequence using the online Graphical Codon Usage Analyzer tool (http://gcua.schoedl.de)[29], we indeed noticed several rare codon usages in the original SBLE gene (Supplementary files: Figure 5). These differences were most striking for arginine (CGC, CGG), glycine (GGC), leucine (CTT) and valine (GTG). Therefore we cloned the codon-optimized construct into the vector of interest and observed a threefold increase in yield compared to the original construct at an induction temperature of 16°C. The protein product from this construct showed a more uniform peak in size-exclusion chromatography. Indeed, under non-optimized conditions, we observed a double peak, and we assumed that the peak eluting first represented a dimeric species. Several lipases are reported to form dimers [30-32]. Our current findings indicated that the first peak is probably the result of a partially unfolded or more extensively glycosylated rSBLE.

Regarding the possible influence of biomass, variation in final optical density (OD_{600nm}) of the different producing strains did not change drastically (Supplementary Figure 6). More precisely, the small variation in cell mass could therefore not be responsible for the observed increases in yield. Indeed, Zhong et al. also observed that, during methanol induction, there was no statistical influence of a lower temperature on the growth curve of recombinant Pichia pastoris[23]. Although their experiment used an induction temperature of 20°C, it is remarkable that these results hold true for 16°C as well.

As a last effort to increase rSBLE expression yield we also investigated the effect of HAC1 on the expression of the codon-optimized construct. Indeed, HAC1 leads to transcriptional activation of endoplasmic-reticulum resident chaperones or foldases, which are normally activated during the UPR pathway [14, 15, 33]. Unfolded protein is typically produced during the high level expression of heterologous proteins, such as rSBLE, and therefore these foldases will recover (partially) unfolded rSBLE, that would normally be degraded [15, 33]. This co-expression led to an even higher increase of yield (Table 3). As of now, we have not yet reached expression levels such as 60 mg/liter for the Geotrichum candidum lipases[34] or, an even higher yield, Rhizomucor miehei lipase (220 mg/liter)[35]. The current lab-scale yield is acceptable for the coming crystallization trials and valorization of this enzyme. In future experiments we will move our expression system to a bioreactor, in order to obtain an even greater increase in yield[36].
**Table 3: Total overview of the different yields during this study.**

<table>
<thead>
<tr>
<th>Expression of:</th>
<th>pPICZαB_rSBLE at 28°C</th>
<th>pPICZαB_rSBLE at 16°C</th>
<th>pPICZαB_rSBLEopt at 16°C</th>
<th>pPICZαB_rSBLEopt + pPICHyg_HAC1spliced at 16°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield (mg/l):</td>
<td>0.75±0.05</td>
<td>4.7±0.5</td>
<td>13.7±1.5</td>
<td>24±1.0</td>
</tr>
<tr>
<td>Improvement</td>
<td>/</td>
<td>Compared to 28°C:</td>
<td>Compared to non-codon-</td>
<td>Compared to no co-expression:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>approximately sixfold</td>
<td>optimized:</td>
<td>approximately twofold</td>
</tr>
</tbody>
</table>

This table represents the total yield ± standard deviation (average of at least 2 technical repeats each). Equal production volumes were compared, as the optical density was not significantly different. After IMAC and size-exclusion purification, the corresponding peak(s) were concentrated to a volume of 1.0 ml and the concentration of the corresponding faction was measured using NanoDrop®.

A remaining bottleneck for crystallization and batch-to-batch reproducibility in protein production is the inherent heterogeneous glycosylation of rSBLE. In previous studies we were able to fully deglycosylate rSBLE *in vitro* using Endo H® without denaturation. Therefore to pave the way for crystallization, the easiest approach to solve this problem would be to perform *in vitro* deglycosylation with Endo H® on the native protein. To counter the batch-to-batch reproducibility in protein production, we also explored co-expression with Endo-N-acetyl-beta-D-glucosaminidase T (EndoT)[37] but this decreased rSBLE expression (results not shown). Although we did not pursue any optimization, a possible explanation could be that complete deglycosylation favors proteolytic degradation or aggregation[38] of rSBLE, therefore, further experiments are needed to assess this. Alternatively, other glyco-engineered strains could be used to express rSBLE[39], for example, using a *P. pastoris* strain, deficient in OCH1p. Although, this strain delivers a more homogeneous N-glycosylation pattern, consisting mainly of Man8GlcNAc2, the yield from such a strain is compromised [40, 41]. Another approach could be to co-express rSBLE with a different Endo-N-acetyl-beta-D-glucosaminidase, such as Endo H.
To conclude, our study solved the two main bottlenecks of rSBLE production in *P. pastoris* and cleared the path for structural and functional studies to understand rSBLE’s illustrious mechanism and, eventually, rSBLE’s valorization.

Acknowledgements

We would like to thank Prof. Dr. Nico Callewaert, for the supply of the *P. pastoris* strain and plasmids.

Compliance with ethical standards

All authors have approved the final version of this article

Funding sources

This work was supported by the European Union's Seventh Framework Program Biosurfing project (PR 289219) and by an Ugent IOF-StarTT (F2014/IOF-StarTT/012) grant; KC and BD are indebted to the support of the ESA-PRODEX project Bioreactor.

Conflict of interest

IVB, KC, BD, and WS have been granted a patent (US9394559 B2) on the enzyme and its use. The other authors declare that they have no conflict of interest.

Ethical approval

This article did not require any studies with human participants or animals by any of the authors.
References


