Journal of Biotechnology 222 (2016) 38–46

Journal of Biotechnology

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

A rapid solubility-optimized screening procedure for recombinant subtilisins in *E. coli*

Gro Elin Kjæreng Bjerga[a,*], Hasan Arsın[b], Øivind Larsen[a], Pål Puntervoll[a], Hans Torstein Kleivdal[a]

[a] *Centre for Applied Biotechnology, Uni Research AS, Thormøhlensgt. 55, N-5008 Bergen, Norway*
[b] *Department of Biology, University of Bergen, Thormøhlensgt. 53 A/B, N-5008 Bergen, Norway*

ARTICLE INFO

Article history:
Received 21 October 2015
Received in revised form 1 February 2016
Accepted 3 February 2016
Available online 6 February 2016

Keywords:
Subtilisin
Serine protease
Recombinant expression
FX-cloning
FITC-casein

ABSTRACT

Subtilisins and other serine proteases are extensively used in the detergent, leather and food industry, and frequently under non-physiological conditions. New proteases with improved performance at extreme temperatures and in the presence of chemical additives may have great economical potential. The increasing availability of genetic sequences from different environments makes homology-based screening an attractive strategy for discovery of new proteases. A prerequisite for large-scale screening of protease-encoding sequences is an efficient screening procedure. We have developed and implemented a screening procedure that encompasses cloning of candidate sequences into multiple expression vectors, cytoplasmic expression in *E. coli*, and a casein-based functional screen. The procedure is plate-format compatible and can be completed in only four days, starting from the gene of interest in a suitable cloning vector. The expression vector suite includes six vectors with combinations of maltose-binding protein (MBP) or the small ubiquitin-related modifier (SUMO) for increased solubility, and polyhistidine tags for downstream purification. We used enhanced green fluorescent protein and four *Bacilli* subtilisins to validate the screening procedure and our results show that proteins were expressed, soluble and active. Interestingly, the highest activities were consistently achieved with either MBP or SUMO fusions, thus demonstrating the merit of including solubility tags. In conclusion, the results demonstrate that our approach can be used to efficiently screen for new subtilisins, and suggest that the approach may also be used to screen for proteins with other activities.

1. Introduction

Proteases are proteolytic enzymes that have great industrial, therapeutic and academic value because of their ability to degrade proteins and peptides (reviewed in Li et al., 2013). In particular, serine proteases have broad applications in the detergent, leather and food industry, due to their broad substrate specificity and their activity at neutral to alkaline pHs (reviewed in Gupta et al., 2002). The industrial conditions under which these proteases are applied may be non-physiological, and include high temperatures, cocktails of detergents and other chemical additives. Discovery and development of proteases that are applicable to industry have a great economic potential, and both sequence- and function-based discovery can be used (examplified in Biver et al., 2013; Kwon et al., 2011). These approaches require a robust production line, which should include molecular cloning and recombinant expression, and functional screening procedures tailored to the relevant proteases (Kwon et al., 2011; Sroga and Dordick, 2002).

Subtilisin-like serine proteases are extensively used in industry, mainly in detergents, and according to a HERA-report,[1] the European Union used about 1000 tons of pure subtilisins in 2002. These proteases are well represented among species of *Bacilli*, are active at an alkaline pH range, and show specificity towards aromatic or hydrophobic residues (Groen et al., 1992). Subtilisins are involved in nutritional regulation in their native hosts and are frequently secreted. They are produced as inactive precursor proteins called pre-pro-proteins or zymogens consisting of a leader sequence that direct their export, a pro-sequence and the catalytic domain. The

* Corresponding author.
E-mail addresses: Gro.Bjerga@uni.no (G.E.K. Bjerga), Hasan.Arsin@uib.no (H. Arsın), Oivind.Larsen@uni.no (Ø. Larsen), Pal.Puntervoll@uni.no (P. Puntervoll), Hans.Kleivdal@uni.no (H.T. Kleivdal).

[1] Human & Environmental Risk Assessment on ingredients of household cleaning products, Edition 2.0 February 2007.

http://dx.doi.org/10.1016/j.jbiotec.2016.02.009

pro-sequence acts as an inhibitor and as a molecular chaperone to guide correct folding of the active enzyme both *in vivo* and *in vitro* (Ikemura et al., 1987; Ohta et al., 1991; Zhu et al., 1989).

Autoproteolytic maturation poses a challenge for heterologous production, but subtilisin-like proteases have been successfully produced in *E. coli* by periplasmic expression (Ikemura and Inouye, 1988; Ikemura et al., 1987). Increased solubility and yield have been achieved with solubility tags, such as the maltose-binding protein (MBP[2]) (Bedouelle and Duplay, 1988; di Guan et al., 1988; Kapust and Waugh, 1999 Kapust and Waugh, 1999), fused to the N-terminus of active serine proteases (Kwon et al., 2011; Sakaguchi et al., 2008). For identification and downstream purification of mature subtilisin-like proteases, they are often expressed in fusion to C-terminal affinity tags, such as polyhistidine (his) (Ghasemi et al., 2012; Hu et al., 2013; Sroga and Dordick, 2002).

In this study, we have developed a screening procedure for rapid and efficient identification of functional subtilisins. The procedure uses *E. coli* as host, implements a versatile cloning system tailored to proteases, followed by recombinant expression in the cytosol, and direct activity screening of lysates using fluorescein isothiocyanate (FITC)-conjugated casein as substrate. The cloning system allows expression of candidate proteases with terminal his-tags for purification, optionally in combination with N-terminal fusions to MBP and the small ubiquitin-related modifier (SUMO) for improved solubility. To validate our approach we have tested four homologous *Bacilli* subtilisins, and show that our screening procedure effectively expresses and detects active, recombinant subtilisins.

2. Material and methods

2.1. Construction of a tailored vector suite

A tailored suite of six different vectors were designed and constructed for the recombinant expression of proteases (Fig. 1, Table 1). p1-p3 vectors contain N-terminal decahistidine tags, and p12, p6 and p7 vectors contain C-terminal hexahistidine tags. The preference for the C-terminal hexahistidine to decahistidine was based on experimental data showing better yields of expression and solubility of eGFP with shorter tag (data not shown). The p2 and p3 vectors were generated by introducing genes encoding the MBP and SUMO fusion partners, respectively, between the existing His-tag and HRV 3C protease (3C) protease site of the p1 vector (Table 1) (Geertsma and Dutzler, 2011). The p12 vector was constructed based on p5 (Table 1) (Geertsma and Dutzler, 2011), where the 3C cleavage site and decahistidine tag was replaced with a hexahistidine tag. Vectors p6 and p7 were constructed by introducing the above-mentioned fusion partners between the start codon and the first SapI site, replacing the 3C and decahistidine region.

For downstream applications, N-terminal affinity and solubility tags were made removable by the utility of specific proteases. A 3C site was designed to p2 and p7, identical to the design in p1 to allow cleavage by 3C protease (Cordingley et al., 1989). No linear motif was added to p3 and p6, as the ubiquitin-like protease 1 (Ulp1) specifically recognizes the tertiary structure of SUMO to remove the tag (Mossessova and Lima, 2000).

Synthetic genes (GenScript) optimized for *E. coli* production encoding either the *E. coli mal*E gene or the *S. cerevisiae smt*3 gene served as templates for megaprimer PCRs of MBP (aa 27–391, GenBank acc. no.: **AIZ93193**) and SUMO (aa 2–97, GenBank acc. no.: **DAA12341**), respectively, by exponential megapriming PCR cloning (EMP, Ulrich et al., 2012). In brief, the megaprimers of genes encoding MBP and SUMO were amplified using Phusion polymerase (NEB), purified with the QIAquick PCR purification kit (Qiagen) and inserted to vectors by linear (in case of p2 and p3) or exponential plasmid amplification (in case of p6 and p7) and treated as described in the protocols (Ulrich et al., 2012; van den Ent and Löwe, 2006). To remove parental DNA, the PCR products were digested with 10U DpnI (NEB) and transformed into *E. coli* MC1061 cells. Plasmids were isolated using the NucleoSpin plasmid purification kit (Macherey-Nagel). Sanger sequencing was used to confirm correct cloning of all vectors. Primers for RF and EMP cloning were designed using an online tool (Bond and Naus, 2012). Information on primers and vectors used in this study is summarized in Table S1 and Table 1, respectively.

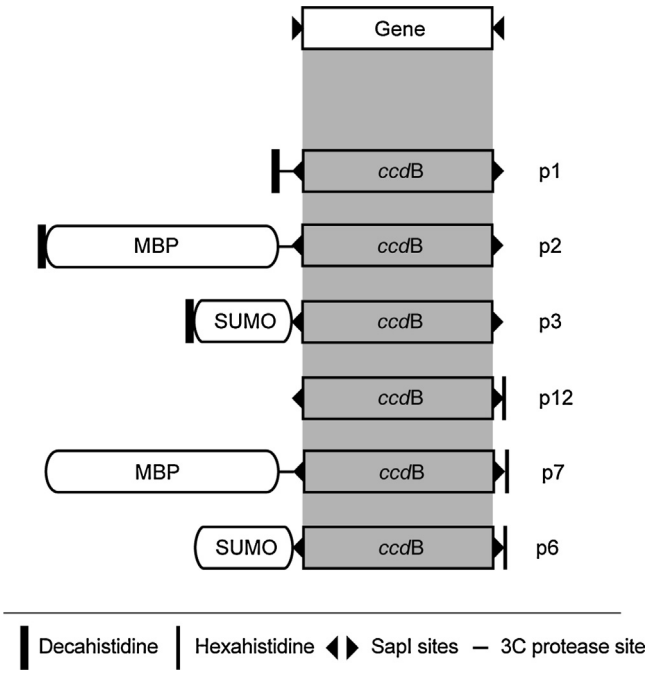

Fig. 1. Schematic representation of the expression vector suite.
Upon restriction-ligation in the FX-cloning regime the gene of interest is replacing the *ccd*B counterselection gene in expression vectors. All vectors, named p1, p2, p3, p12, p7 and p6, contain his-tags, either N-terminally (decahistidine) or C-terminally (hexahistidine). In addition, four of the vectors contain the MBP and SUMO solubility partners as fusions to the N-terminies of the protease sequences. Triangles show placement and orientation of the SapI sites used in the FX-cloning procedure. Linear sequence motifs are introduced to allow tag removal by 3C protease in p1, p2 and p7 (black line), whereas the SUMO fusion in p3 and p6 can be removed by Ulp1 protease after tertiary structure recognition.

2.2. Molecular cloning of eGFP and subtilisins to vector suite

The pCMV cyto-EGFP-myc plasmid served as template for the amplification of a mutated *gfp* gene (*egfp)* encoding enhanced green fluorescent protein (eGFP, residues 2-239) (Cormack et al., 1996) by Phusion PCR and primers in Table S1. The *egfp* gene was integrated into the pINITIAL cloning vector (Table 1) by digesting the PCR product and the vector with SapI (NEB) and ligating with T4 DNA ligase (NEB) according to the fragment exchange (FX) cloning protocol (Geertsma, 2014). The ligation reaction was transformed into *E. coli* MC1061 cells and clones were selected on LB-agar (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 1.5% (w/v) agar-agar) supplemented with kanamycin (50 μg/ml, Sigma-Aldrich). Plasmids were isolated as above, and sequencing was used to confirm correct cloning of all pINITIAL constructs. Sub-cloning from pINITIAL to the expression vectors (Fig. 1) were carried out as outlined

[2] Abbreviations: his, polyhistidine; MBP, maltose binding protein, SUMO, small ubiquitin-related modifier; FX-cloning, fragment exchange cloning; RF, restriction-free cloning; EMP, exponential megapriming PCR; 3C, HRV 3C protease; Ulp1, ubiquitin-like protease 1, FITC, fluorescein isothiocyanate; HRP, horseradish peroxidase; eGFP, enhanced green fluorescent protein.

Table 1
Vectors used in this study.

Vector name	Fusion partners	His-tag length (aa)	Fusion partner size (kDa)	Promoter	Tag removal[a]	Resistance gene	Vector type	Reference
pINITIAL	–	–	–	–	–	*cam*	Cloning	Geertsma and Dutzler (2011)
p1 (pBXNH3)	N-his	10	2.6	*pBAD*	3C	*amp*	Expression	Geertsma and Dutzler (2011)
p2	N-his-MBP	10	42.7	*pBAD*	3C	*amp*	Expression	This study
p3	N-his-SUMO	10	12.6	*pBAD*	Ulp1	*amp*	Expression	This study
p5 (pBXC3H)	C-his	10	2.3	*pBAD*	3C	*amp*	Expression	Geertsma and Dutzler (2011)
p12	C-his	6	0.8	*pBAD*	–	*amp*	Expression	This study
p7	N-MBP, C-his	6	41.4	*pBAD*	3C	*amp*	Expression	This study
p6	N-SUMO, C-his	6	11.3	*pBAD*	Ulp1	*amp*	Expression	This study
pUC57:*smt*3 (S2-G97)-opt	(*smt*3 template)	–	–	–	–	*kan*	Cloning	This study
pUC57:His-*mal*E(K27-Q391) 3C-opt	(*mal*E template)	–	–	–	–	*kan*	Cloning	This study
pUC57.kan, SapI-free	–	–	–	–	–	*kan*	Cloning	GenScript

[a] For p6 and p7 the N-terminal fusion partners, but not the C-terminal his tags, are removable.

above, except that clones were selected on LB agar with ampicillin (100 μg/ml, Sigma-Aldrich).

Codon-optimized *apr* genes (GenScript) encoding *Bacilli* subtilisins (Table 2) flanked by SapI sites served as templates for FX cloning of full-length genes to expression vectors, or served as templates for PCRs in the case of truncated genes. Otherwise, genes were cloned as described above. Empty vectors were generated by replacing the *ccd*B gene with a GSGSGS linker to allow their cloning and expression in *E. coli* MC1061 cells. The GSGSGS-linker was constructed by hybridizing two oligos to one double stranded DNA fragment designed to contain sticky SapI overhangs to allow its integration to the pINITIAL vector. Information on primers used is summarized in Table S1.

2.3. Site-directed mutagenesis

A Phusion PCR using mutagenesis primers (Table S1) was designed to generate a S325A mutation in the *apr* gene of *B. licheniformis* DSM13 in pINITIAL. Parental DNA was removed by DpnI digestion. Mutants were then transformed into *E. coli* MC1061 and selected to LB agar containing 34 μg/ml chloramphenicol (Sigma-Aldrich). The mutant subtilisin was sub-cloned to the expression vector suite as described above.

2.4. Recombinant expression

Recombinant expression was carried out according to a protocol described elsewhere (Vincentelli et al., 2011). The *E. coli* MC1061 strain was utilized for expression due to its inability to metabolize the inducer, L-arabinose. Precultures in LB media were inoculated directly from positive clones on LB agar with 100 μg/ml ampicillin and grown 16–20 h at 250 rpm at 37 °C. In deep 24-well plates, 4 ml 2YT media (1.6% (w/v) tryptone, 1% (w/v) yeast extract and 0.5% (w/v) NaCl) containing 100 μg/ml ampicillin was inoculated with 100 μl precultures and incubated at 37 °C at 250 rpm for 2 h to reach log phase. Cultures were equilibrated to 20 °C for 30 min before induction of the *pBAD* promoter with 0.1% (v/v) L-arabinose (Sigma-Aldrich) for 16–20 h at 250 rpm at 20 °C. To improve the protocol, cultures can also be autoinduced by replacing lactose with arabinose in a trace-metal free version of the ZYP-5052 media (Studier, 2005). 100 μl culture was used for OD measurements by reading absorbance at 600 nm using the Hidex Sense microplate reader (Kem-En-Tec Nordic). The cells were harvested using an Allegra X-12R benchtop centrifuge (Beckman Coulter) at 4750 rpm for 15 min. Cells were resuspended in 1 ml 8.5 N lysis buffer (50 mM Tris HCl pH 8.5, 50 mM NaCl, 0.25 mg/ml lysozyme, 10% (v/v) glycerol), and incubated at 20 min at room temperature during gentle agitation for lysis. Lysis was completed by ultrasound using five seconds pulse two times at 40–60% amplitude with a CV-18 probe powered by an Ultrasonic Homogenizer 4710 (Cole Parmer). Lysates were cleared by centrifugation at 4750 rpm for 15 min. Cleared lysate samples (representing soluble fraction) were analysed by SDS-PAGE (Laemmli, 1970).

2.5. Immunoblot

Proteins from cleared lysates were analysed by SDS-PAGE, and transferred onto a nitrocellulose membrane (Towbin et al., 1979)

Table 2
Subtilisins used in this study.

Origin (short name)	Enzyme name	GenBank acc. no.	ATCC reference	Reference (protease)	Reference (genome)	Length (aa)	Sequence identity to *B. licheniformis* DSM13 (%).
Bacillus licheniformis DSM13 (*Bli*)	Subtilisin Carlsberg AprE	**AAU40017**	14580	Jacobs et al. (1985)	Rey et al. (2004)	379	100
Bacillus paralicheniformis ATCC 9945a (*Bpa*)	Subtilisin Carlsberg AprE	**AGN35600**	9945a	Jacobs et al. (1985)	Rachinger et al. (2013)	379	98.2
Bacillus subtilis subsp. subtilis str. 168 (*Bsu*)	SubtilisinE	**CAB12870**	23857	Stahl and Ferrari (1984)	Kunst et al. (1997)	381	66.1
Bacillus amyloliquefaciens ATCC 23844 (*Bam*)	BPN' subtilisin	**AAB05345**	23844	Vasantha et al. (1984)	–	382	66.1

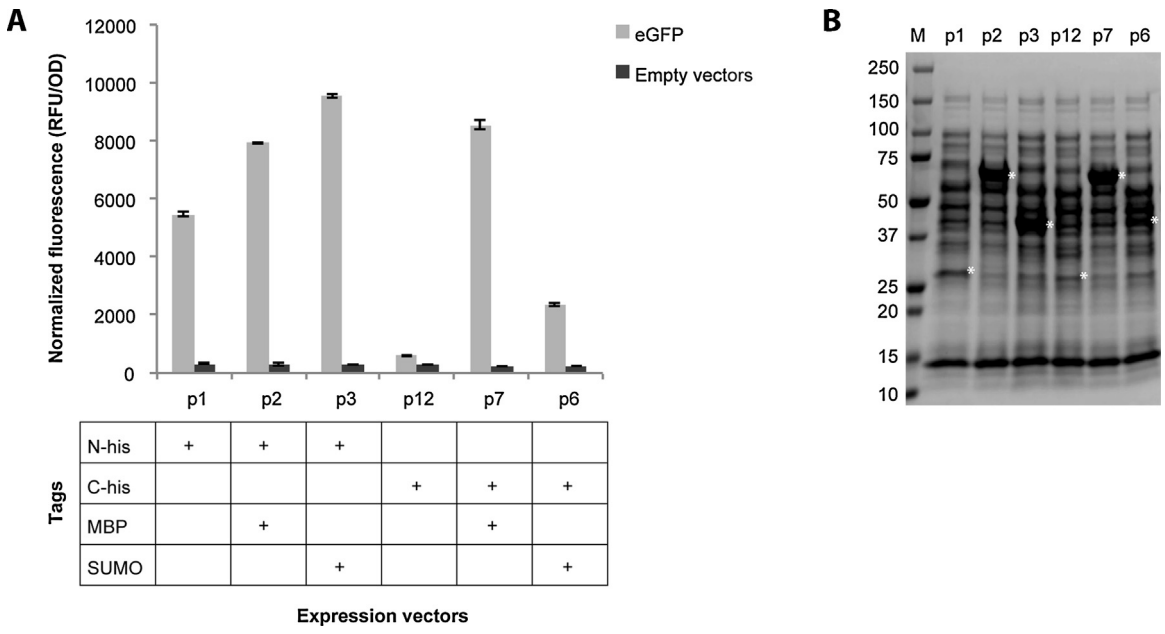

Fig. 2. Recombinant expression of eGFP in the vector suite.
A. Fluorescence from cultures containing eGFP in six different fusion constructs. Fluorescence was normalized to the optical density of the bacterial cultures. Expression from empty vectors was used as a control. The Table below shows the presence of tags in the different constructs. B. Analysis of protein integrity of soluble eGFP fusion proteins by SDS-PAGE. Sizes of molecular weight standard (M) are shown to the right (in kDa). Asterisks indicate the expected mass of the recombinant proteins.

using the Trans-Blot Turbo (BioRad) transfer system. A mouse monoclonal anti-polyhistidine antibody (H1029, Sigma-Aldrich) and a mouse monoclonal anti-MBP antibody (M1321, Sigma-Aldrich) were used to detect recombinant expression of subtilisin mutants. The primary antibodies were detected with a secondary rabbit HRP-linked mouse IgG (NA931 V, GE Healtcare). The HRP-reaction was developed with the Clarity Western ECL Substrate (BioRad), and imaged in the Chemi-Doc gel imager (BioRad).

2.6. eGFP fluorescence measurement

Upon harvest, 100 μl culture from eGFP expression was used for OD measurements by reading the absorbance at 600 nm using the FluoStar Optima microplate reader (BMG Labtech). Fluorescence from cell cultures was measured at excitation 485 nm and emission 520 nm, and values were normalized to optical density of the bacterial cultures to compensate for growth differences.

2.7. Casein-based protease assay

The protease fluorescent detection kit (PF0100, Sigma-Aldrich) was used for detection of proteolytic activity (Twining, 1984). 5 μl cleared lysates were used in activity assays with 10 μl FITC-labelled casein and adjusted to 25 μl with 20 mM phosphate buffer pH 7.6, and incubated at 60 °C for 1 h. Unhydrolyzed proteins were precipitated by 75 μl 0.6 N Trichloroacetic Acid (TCA) at 37 °C for 30 min. Hydrolyzed FITC-peptides were then collected by centrifugation in a Heraeus Pico centrifuge (Thermo Scientific) at 13.000 rpm for 5–10 min. 2 μl supernatant was added to 200 μl 0.5 M TrisHCl pH 8.5 in MicroFluor 1 plates (Thermo Scientific) and fluorescence was measured at excitation 485 nm and emission 520 nm using the Hidex Sense microplate reader. As a control, Alcalase ®2.4 L (P4860, Sigma-Aldrich) was used at a dilution 1:50 000 in lysis buffer.

3. Results and discussion

3.1. Screening procedure design and vector suite validation

Our aim was to develop a quick but robust screening procedure for subtilisin-like proteases. We identified a set of design requirements for such a screening procedure: the entire procedure should be compatible with a plate format for high throughput; expression should be carried out in the cytoplasm of *E. coli* for high yield; the cloning system should facilitate simple, parallel sub-cloning into a set of vectors for exploitation of purification and solubility tags; and finally, a functional screen should be implemented for assessment of activity from recombinant proteases in lysates. A recently developed vector suite, which facilitates sub-cloning based on fragment exchange (FX) of the gene of interest from a cloning vector into multiple expression vectors, was chosen as a starting point (Geertsma and Dutzler, 2011). The approach is highly effective due to the directional cloning caused by the orientation of the type IIS restriction sites (SapI) and the presence of counter selection genes. Based on available vectors (p1 and p5, Table 1), we constructed five new ones, p2, p3, p6, p7 and p12 (Fig. 1, Table 1), designed to express serine proteases in fusion to selected affinity and solubility tags. We chose to include two solubility tags for use in N-terminal fusions, namely MBP (Kapust and Waugh, 1999) and SUMO (Malakhov et al., 2004). In case the tags would interfere with maturation and/or proteolytic activity, specific protease sites were included for tag removal (3C and Ulp1 proteases, respectively). All constructs contained either an N-terminal or C-terminal his-tag to enable detection and protein purification. The C-terminal his-tag containing vectors were constructed to allow purification of mature protease (Ghasemi et al., 2012; Hu et al., 2013; Sroga and Dordick, 2002). Cloning was scaled to a plate-based format that allowed parallel cloning of many constructs simultaneously.

To validate the expression vectors, we chose the gene encoding eGFP for simple fluorescence-based monitoring of recombinant protein production. The *egfp* gene was cloned into the pINITIAL cloning vector, verified by sequencing, and sub-cloned into our

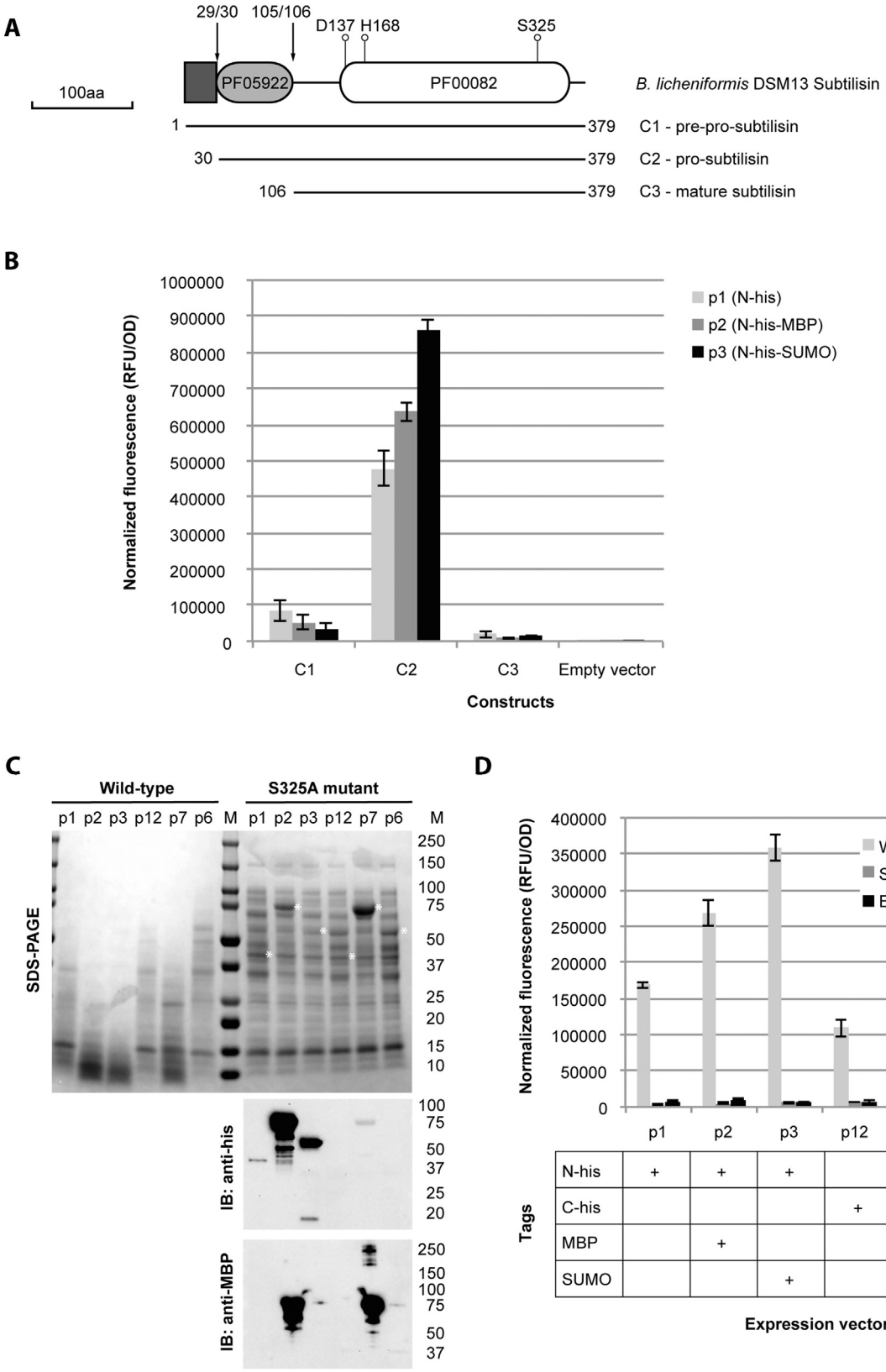

Fig. 3. Activity of recombinant *B. licheniformis* DSM13 pro-subtilisin in *E. coli.*

A. Cartoon of the annotated *B. licheniformis* DSM13 subtilisin with leader sequence (square-shaped box), proteolytic cleavage positions (arrows), catalytic triad (open rings) and Pfam domain annotations (PF-coded oval-shaped box). Below the constructs used in this study are shown: the full-length pre-pro subtilisin sequence (residues 1-379, C1), the pro-subtilisin construct (residues 30-379, C2), and the mature subtilisin (residues 106-379, C3). Cartoon is drawn to scale. B. The soluble fraction of extracts containing subtilisin constructs (C1-C3) in three vectors, p1, p2 or p3, were screened for proteolytic activity against FITC-conjugated casein. Fluorescence was normalized to optical density of expression cultures, to account for growth effects. Expression from empty vectors was used as a control. Error bars represent standard deviation between two independent experiments. C. Analysis of protein integrity of native pro-subtilisin and the catalytic S325A mutant by SDS-PAGE (upper panel) and immunoblots (IB). Antibodies against the his-tag (middle panel) and the MBP tag (lower panel)

panel of expression vectors. Next, the vectors were used to express the eGFP fusion proteins in *E. coli* MC1061, and the expression cultures were monitored directly for *in vivo* fluorescence (Fig. 2A) (such as applied in Scholz et al., 2013). The N-terminally his-tagged eGFP construct (p1) had a much higher level of fluorescence than the one with a C-terminal his-tag (p12), which fluoresced just above background. Both the N- and C-terminally his-tagged constructs displayed increased fluorescence when fused to either MBP or SUMO (Fig. 2A). Protein expression levels in the soluble fraction correlated well with the *in vivo* fluorescence data (Fig. 2B), suggesting that the differences in fluorescence were due to different amounts of soluble protein. These results validated the integrity of the vectors, and demonstrated that the MBP and SUMO tags increased the solubility of the target protein, in line with previous data (Marblestone et al., 2006).

3.2. *Validation of the screening procedure using four* Bacillus *subtilisins*

To validate the suitability of the vector suite for expression of serine proteases, the industrially relevant subtilisin from *Bacillus licheniformis* (trade name: Alcalase 2.4 L, Novozymes™) was chosen as a reference protease. As periplasmic expression of subtilisin in *E. coli* has shown low expression levels (Ikemura and Inouye, 1988; Ikemura et al., 1987), we decided to attempt cytoplasmic expression (e.g. Hu et al., 2013; Maciver et al., 1994; Zhang et al., 2005).

To investigate whether expressing the full-length or other truncated versions of the *apr* gene would affect solubility and activity, we cloned three versions of the gene of the *Bacillus licheniformis* DSM13 *apr* into the p1, p2, and p3 vectors (Fig. 3A). The first construct contained the entire *apr* gene, encoding pre-pro subtilisin (residues 1-379, C1). The second construct contained a truncated gene encoding pro-subtilisin (residues 30-379, C2) without the leader sequence. The third construct contained a truncated gene where the part encoding both the pre and pro sequences was removed (residues 106-379, C3). All gene versions were codon-optimized to improve protein expression. The three *apr* gene variants were cloned into pINITIAL, sub-cloned into the three expression vectors, and expressed in the *E. coli* MC1061 strain.

To measure the activity of the expressed subtilisin fusions, we chose an *in vitro* casein-based protease assay (Twining, 1984). In this assay, the use of fluorescein isothiocyanate (FITC) conjugated casein allows direct screens for protease activity in the cleared extracts, thus eliminating elaborate purification steps. Casein was chosen to facilitate screening for proteases with broad specificity for peptide bonds. The soluble extracts containing subtilisin fusions were tested in the assay, and compared to extracts from strains carrying empty plasmid vectors (Fig. 3B). The negative controls using empty vectors did not exhibit any activity, which suggests that the activity observed in the extracts containing subtilisin fusion was not due to *E. coli* host proteases. Hence, the proteolytic degradation of FITC-casein in those extracts is likely caused by the activity of the expressed native subtilisin fusions. Cultures expressing the C2 version were shown to be the optimal construct (Fig. 3B). The C1 version showed significant growth deficiencies, which may suggest that the subtilisin leader sequence caused misfolding. Despite this, some activity was observed with soluble extracts containing C1 in the FITC-casein assay. In contrast, soluble extracts containing the C3 construct showed no activity. In line with previous data, these results strongly suggest that putative leader sequences should be removed when screening for new subtilisins using our system, but that the pro sequence should be kept intact (Ikemura et al., 1987; Ohta et al., 1991; Zhu et al., 1989).

It was not obvious that N-terminal fusion tags would promote correct folding of subtilisin, as these might hinder folding assisted by the pro domain. The highest activity was, however, observed with the extracts containing the N-terminal his-SUMO (p2) and his-MBP (p3) fusion constructs of the C2 version of subtilisin (Fig. 3B). These results show that fusion to a large solubility partner at the N-terminal of pro-subtilisin promoted production of active serine proteases. Our data is in line with previous data obtained with MBP (Kwon et al., 2011; Sakaguchi et al., 2008). To our knowledge, this is the first time subtilisin has been successfully expressed as a SUMO fusion protein in *E. coli*.

As a C-terminal his-tag may allow downstream purification and detection after maturation, the C2 subtilisin construct was also cloned to the p6, p7 and p12 vectors. As an additional negative control, we constructed a mutant of the truncated *apr* gene where the codon for the nucleophilic serine S325 in the catalytic triad was mutated to an alanine codon (Fig. 1) (Carter and Wells, 1988). The soluble fractions of sonicated cells were analysed by SDS-PAGE, revealing striking differences in the observed protein patterns of the native subtilisin fusion proteins and their cognate mutant controls (Fig. 3C). The lack of high molecular weight proteins in the native subtilisin-containing fractions suggests that they are expressed and have proteolytic activity. However, there were no visible protein bands corresponding to the native subtilisin fusion proteins, and mass spectrometry (MS) analysis of the samples also failed to identify them. Instead, the MS analysis revealed the presence of known protease resistant *E. coli* membrane and periplasmic proteins (Table S2). In contrast to the native subtilisin fusion proteins, most of the mutant subtilisins, with the exception of the p12 mutant construct, were identified, either directly on the Coomassie-stained SDS-PAGE, or by immunoblotting using anti-his and anti-MBP antibodies (Fig. 3C). Overall, this suggests that the subtilisin fusion proteins are expressed, soluble and active.

All subtilisin fusion proteins were able to digest FITC-casein (Fig. 3D), but again the highest activity was observed with the extracts containing the N-terminal his-SUMO and his-MBP fusion constructs. Interestingly, the conventional subtilisin construct design with a C-terminal his-tag (p12) was the second lowest performing construct in the activity assay. The extent of hydrolysis observed by the SDS-PAGE analysis (Fig. 3C) correlated with the level of proteolytic activity observed in the FITC-casein assay (Fig. 3D).

Furthermore, there is a striking similarity to the eGFP validation experiment (Fig. 2); the effect of the different fusion constructs on the levels of fluorescence in the eGFP experiment is similar to the activities observed with the subtilisin fusion proteins (Fig. 3D). This suggests that the vector suite may have some generic features, and that its utility is not restricted to proteases. To conclude, the described *E. coli* based expression system was used to successfully express soluble *B. licheniformis* DSM13 subtilisin fusion proteins, which were active in the casein-based activity assay.

As a final validation step of our screening procedure, we performed a mini screen of the *B. licheniformis* DSM13 *apr* gene and three other *Bacilli genes* from *B. paralicheniformis* ATCC 9945a, *B. amyloliquefaciens* ATCC 23844 and *B. subtilis* subsp. *subtilis* str. 168 (Table 2). The subtilisins encoded by these genes are thermophilic

were used to identify the recombinant mutant constructs. Sizes of molecular weight standard (M) are shown to the right (in kDa). Asterisks indicate the expected mass of the unprocessed recombinant proteins. The mature subtilisin is 27 kDa D. FITC-casein activity of soluble *E. coli* extracts containing fusion constructs of native pro-subtilisin from *B. licheniformis* DSM13 and the catalytic S325A mutant. Hydrolysis was measured as in B, except error bars show standard deviation in two replicates in one representative experiment.

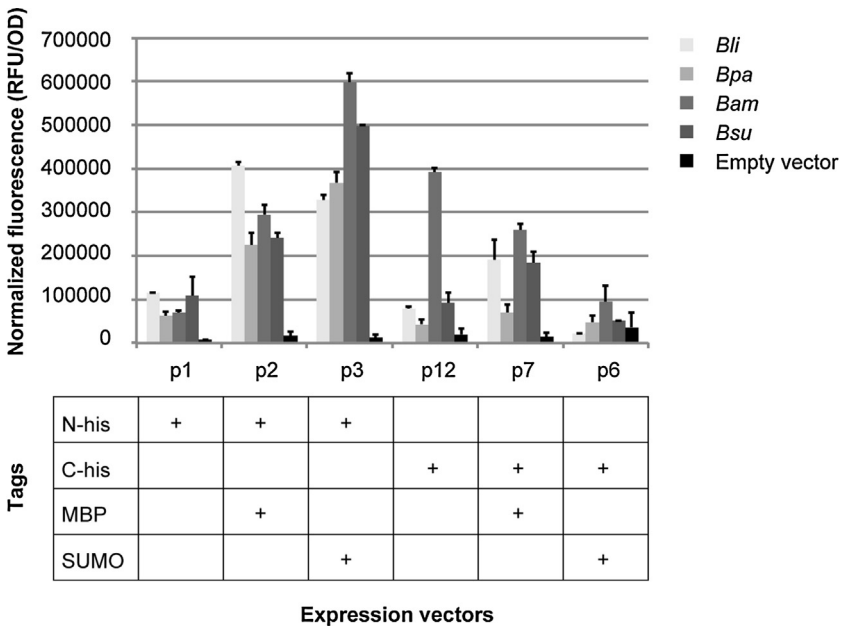

Fig. 4. Proteolytic activity of four *Bacilli* subtilisin homologues.
The soluble fractions of *E. coli* extracts containing fusion constructs of the four homologous subtilisins were screened for proteolytic activity against FITC-conjugated casein. Hydrolysis was measured as in Fig. 3, where error bars show standard deviation in two replicates in one representative experiment. *Bli, Bacillus licheniformis* DSM13*; Bpa, Bacillus paralicheniformis* ATCC 9945a*; Bsu, Bacillus subtilis* subsp. *subtilis* str. 168*; Bam, Bacillus amyloliquefaciens* ATCC 23844.

with temperature optimums above 40 °C (Table 2) (examples in Ferreira et al., 2003; Peng et al., 2003; Sellami-Kamoun et al., 2008), and their pairwise sequence identities range from 66 to 98% (Table 2). The three additional *apr* genes, without the leader sequence encoding part, were cloned into the six expression vectors. Each of the four subtilisins was expressed from the six expression plasmids, and soluble extracts were tested for protease activity using the FITC-casein assay (Fig. 4). All four subtilisin targets were positively identified as active proteases in four or five of the constructs, using three times the background signal to determine the threshold. The two constructs that showed the most consistent activity across the different subtilisins were the his-MBP (p2) and his-SUMO fusions (p3). This suggests that one could restrict future screens for new subtilisins to only the p2 and p3 vectors, thus increasing the throughput of the system. However, the high activity of the p12 construct of *B. amyloliquefaciens* ATCC 23844 subtilisin (Fig. 4) demonstrates the merit of including additional vectors in the screen, as the C-terminal his tag allows for downstream purification and characterization. This construct also indicates that the success of recombinant expression from a particular vector may be protein-dependent, and that a range of expression vehicles with different properties may be valuable to identify positive candidates.

4. Conclusion

The development of a screening procedure for the identification of recombinant subtilisins has been described. The overall procedure enables a four-day approach from the sub-cloning of candidate genes in expression vectors to the evaluation of enzymatic activity (Fig. 5). The approach utilizes a rapid cloning method for sub-cloning of genes into six expression vectors and subsequent recombinant expression in *E. coli* MC1061, an ultrasound-based cell lysis and direct activity assessment of soluble fractions with FITC-casein. The screening approach has been validated using four homologous *Bacilli apr* encoded subtilisins of varying degree of sequence identities (ranging from 65 to 98%) and temperature profiles (ranging roughly 40–80 °C). All subtilisins were identified as active proteases in at least four of the constructs. Hence, this approach is suitable for application in discovery of novel subtilisin-like proteases or for application on engineered subtilisin-like proteases (Bryan, 2000; Wells and Estell, 1988) to effectively screen production of mutants and their enzymatic activity. We also showed, using eGFP, that the utility of the vector suite is not restricted to proteases. Other applications may, however, require tailoring to the specific type of protein, depending on their properties, such as metal-requirement, temperature and pH range.

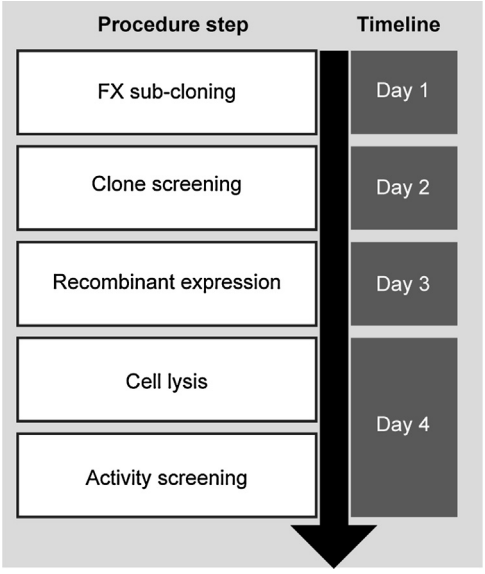

Fig. 5. A flow-chart of the rapid, solubility-optimized screening procedure for recombinant subtilisins in *E. coli*.
The procedure has been developed and tailored to subtilisins, and enables a four-day approach succeeding from sub-cloning of candidate genes in expression vectors to the evaluation of enzymatic activity. The screening-approach utilizes a rapid and versatile FX-cloning regime, subsequent recombinant production in *E. coli* MC1061, an ultrasound-based cell lysis and activity assaying with FITC-casein.

Contributions

GEKB, PP, ØL, HTK jointly conceived and designed the study, GEKB has performed the bioinformatical analyses, GEKB and HA carried out the experiments and supervised students, GEKB and PP prepared the manuscript. All authors have discussed the results and read, edited and approved the final manuscript.

Acknowledgements

We would like to thank the students Odin Blomset and Anni Geithus for excellent technical assistance. The pCMV-cyto-EGFP-myc plasmid was kindly provided by Prof. Mathias Ziegler's lab, University of Bergen. This study was financed by the Norwegian Research Council as part of the NorZymeD initiative (project ID: 221568).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbiotec.2016.02.009.

References

Bedouelle, H., Duplay, P., 1988. Production in *Escherichia coli* and one-step purification of bifunctional hybrid proteins which bind maltose: export of the Klenow polymerase into the periplasmic space. Eur. J. Biochem. 171, 541–549.

Biver, S., Portetelle, D., Vandenbol, M., 2013. Characterization of a new oxidant-stable serine protease isolated by functional metagenomics. Springerplus 2, 410, http://dx.doi.org/10.1186/2193-1801-2-410.

Bond, S.R., Naus, C.C., 2012. RF-Cloning.org: an online tool for the design of restriction-free cloning projects. Nucleic Acids Res. 40, W209–13, http://dx.doi.org/10.1093/nar/gks396.

Bryan, P.N., 2000. Protein engineering of subtilisin. Biochim. Biophys. Acta 1543, 203–222, http://dx.doi.org/10.1016/S0167-4838(00)00235-1.

Carter, P., Wells, J.A., 1988. Dissecting the catalytic triad of a serine protease. Nature 332, 564–568, http://dx.doi.org/10.1038/332564a0.

Cordingley, M.G., Register, R.B., Callahan, P.L., Garsky, V.M., Colonno, R.J., 1989. Cleavage of small peptides in vitro by human rhinovirus 14 3C protease expressed in Escherichia coli. J. Virol. 63, 5037–5045.

Cormack, B.P., Valdivia, R.H., Falkow, S., 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173, 33–38, http://dx.doi.org/10.1016/0378-1119(95)00685-0.

di Guan, C., Li, P., Riggs, P.D., Inouye, H., 1988. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. Gene 67, 21–30, http://dx.doi.org/10.1016/0378-1119(88)90004-2.

Ferreira, L., Ramos, M.A., Dordick, J.S., Gil, M.H., 2003. Influence of different silica derivatives in the immobilization and stabilization of a Bacillus licheniformis protease (Subtilisin Carlsberg). J. Mol. Catal. B Enzym. 21, 189–199, http://dx.doi.org/10.1016/S1381-1177(02)00223-0.

Geertsma, E.R., Dutzler, R., 2011. A versatile and efficient high-throughput cloning tool for structural biology. Biochemistry 50, 3272–3278, http://dx.doi.org/10.1021/bi200178z.

Geertsma, E.R., 2014. FX cloning: a simple and robust high-Throughput cloning method for protein expression. In: Valla, S., Lale, R. (Eds.), DNA Cloning and Assembly Methods, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 153–164, http://dx.doi.org/10.1007/978-1-62703-764-8.

Ghasemi, Y., Dabbagh, F., Ghasemian, A., 2012. Cloning of a fibrinolytic enzyme (subtilisin) gene from *Bacillus subtilis* in *Escherichia coli*. Mol. Biotechnol. 52, 1–7, http://dx.doi.org/10.1007/s12033-011-9467-6.

Groen, H., Meldal, M., Breddam, K., 1992. Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching. Biochemistry 31, 6011–6018, http://dx.doi.org/10.1021/bi00141a008.

Gupta, R., Beg, Q.K., Lorenz, P., 2002. Bacterial alkaline proteases: molecular approaches and industrial applications. Appl. Microbiol. Biotechnol. 59, 15–32, http://dx.doi.org/10.1007/s00253-002-0975-y.

Hu, H., He, J., Yu, B., Zheng, P., Huang, Z., Mao, X., Yu, J., Han, G., Chen, D., 2013. Expression of a keratinase (kerA) gene from *Bacillus licheniformis* in *Escherichia coli* and characterization of the recombinant enzymes. Biotechnol. Lett. 35, 239–244, http://dx.doi.org/10.1007/s10529-012-1064-7.

Ikemura, H., Inouye, M., 1988. In vitro processing of pro-subtilisin produced in *Escherichia coli*. J. Biol. Chem. 263, 12959–12963.

Ikemura, H., Takagi, H., Inouye, M., 1987. Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli*. J. Biol. Chem. 262, 7859–7864.

Jacobs, M., Eliasson, M., Uhlén, M., Flock, J.I., 1985. Cloning: sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. Nucleic Acids Res. 13, 8913–8926.

Kapust, R.B., Waugh, D.S., 1999. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 8, 1668–1674, http://dx.doi.org/10.1110/ps.8.8.1668.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, a M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, a, Borchert, S., Borriss, R., Boursier, L., Brans, a, Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C., Caldwell, V., Capuano, B., Carter, V., Choi, N.M., Codani, S.K., Connerton, J.J., Danchin, I.F., 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390, 249–256, http://dx.doi.org/10.1038/36786.

Kwon, K., Hasseman, J., Latham, S., Grose, C., Do, Y., Fleischmann, R.D., Pieper, R., Peterson, S.N., 2011. Recombinant expression and functional analysis of proteases from *Streptococcus pneumoniae*, *Bacillus anthracis*, and *Yersinia pestis*. BMC Biochem. 12, 17, http://dx.doi.org/10.1186/1471-2091-12-17.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685, http://dx.doi.org/10.1038/227680a0.

Li, Q., Yi, L., Marek, P., Iverson, B.L., 2013. Commercial proteases: present and future. FEBS Lett. 587, 1155–1163, http://dx.doi.org/10.1016/j.febslet.2012.12.019.

Maciver B., Mchale R.H., Saul D.J., Bergquist P.L., 1994, Cloning and sequencing of a serine proteinase gene from a thermophilic *Bacillus* species and its expression in *Escherichia coli*. Cloning and Sequencing of a Serine Proteinase Gene from a Thermophilic *Bacillus* Species and Its Expression in *Escherichia coli* 60.

Malakhov, M.P., Mattern, M.R., Malakhova, O.A., Drinker, M., Weeks, S.D., Butt, T.R., 2004. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J. Struct. Funct. Genomics 5, 75–86, http://dx.doi.org/10.1023/B:JSFG.0000029237.70316.52.

Marblestone J.G., Edavettal S.C., Lim Y., Lim P., Zuo X.U.N., Butt T.R., 2006, Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO 182–189. 10.1110/ps.051812706.for.

Mossessova, E., Lima, C.D., 2000. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol. Cell 5, 865–876.

Ohta, Y., Hojo, H., Aimoto, S., Kobayashi, T., Zhu, X., Jordan, F., Inouye, M., 1991. Pro-peptide as an intramolecular chaperone: renaturation of denatured subtilisin E with a synthetic pro-peptide [corrected]. Mol. Microbiol. 5, 1507–1510, http://dx.doi.org/10.1111/j.1365-2958.1991.tb00797.x.

Peng, Y., Huang, Q., Zhang, R.H., Zhang, Y.Z., 2003. Purification and characterization of a fibrinolytic enzyme produced by *Bacillus* amyloliquefaciens DC-4 screened from douchi, a traditional Chinese soybean food. Comp. Biochem. Physiol.—B Biochem. Mol. Biol. 134, 45–52, http://dx.doi.org/10.1016/S1096-4959(02) 00183-5.

Rachinger, M., Volland, S., Meinhardt, F., Daniel, R., Liesegang, H., 2013. First insights into the completely annotated genome sequence of bacillus licheniformis strain 9945A. Genome Announc. 1, 525–526, http://dx.doi.org/10.1128/genomeA.00525-13.

Rey, M.W., Ramaiya, P., Nelson, B.A., Brody-Karpin, S.D., Zaretsky, E.J., Tang, M., Lopez de Leon, A., Xiang, H., Gusti, V., Clausen, I.G., Olsen, P.B., Rasmussen, M.D., Andersen, J.T., Jørgensen, P.L., Larsen, T.S., Sorokin, A., Bolotin, A., Lapidus, A., Galleron, N., Ehrlich, S.D., Berka, R.M., 2004. Complete genome sequence of the industrial bacterium Bacillus licheniformis and comparisons with closely related Bacillus species. Genome Biol. 5, R77, http://dx.doi.org/10.1186/gb-2004-5-10-r77.

Sakaguchi, M., Niimiya, K., Takezawa, M., Toki, T., Sugahara, Y., Kawakita, M., 2008. Construction of an expression system for aqualysin I in *Escherichia coli* that gives a markedly improved yield of the enzyme protein. Biosci. Biotechnol. Biochem. 72, 2012–2018, http://dx.doi.org/10.1271/bbb.80132.

Scholz, J., Besir, H., Strasser, C., Suppmann, S., 2013. A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. BMC Biotechnol. 13, 12, http://dx.doi.org/10.1186/1472-6750-13-12.

Sellami-Kamoun, A., Haddar, A., Ali, N.E.H., Ghorbel-Frikha, B., Kanoun, S., Nasri, M., 2008. Stability of thermostable alkaline protease from *Bacillus licheniformis* RP1 in commercial solid laundry detergent formulations. Microbiol. Res. 163, 299–306, http://dx.doi.org/10.1016/j.micres.2006.06.001.

Sroga, G.E., Dordick, J.S., 2002. A strategy for in vivo screening of subtilisin E reaction specificity in *E. coli* periplasm. Biotechnol. Bioeng. 78, 761–769, http://dx.doi.org/10.1002/bit.10269.

Stahl, M.L., Ferrari, E., 1984. Replacement of the *Bacillus subtilis* subtilisin structural gene with an in vitro-derived deletion mutation. J. Bacteriol. 158, 411–418.

Studier, F.W., 2005. Protein production by Auto-induction in high-density shaking cultures. Protein Expr. Purif. 41, 207–234, http://dx.doi.org/10.1016/j.pep.2005.01.016.

Towbin, H., Staehelin, T., Gordon, J., 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. U. S. A. 76, 4350–4354, http://dx.doi.org/10.1073/pnas.76.9.4350.

Twining, S.S., 1984. Fluorescein isothiocyanate-labeled casein assay for proteolytic enzymes. Anal. Biochem. 143, 30–34, http://dx.doi.org/10.1016/0003-2697(84) 90553-0.

Ulrich, A., Andersen, K.R., Schwartz, T.U., 2012. Exponential megapriming PCR (EMP) cloning–seamless DNA insertion into any target plasmid without sequence constraints. PLoS One 7, e53360, http://dx.doi.org/10.1371/journal.pone.0053360.

Vasantha, N., Thompson, L.D., Rhodes, C., Banner, C., Nagle, J., Filpula, D., 1984. Genes for alkaline protease and neutral protease from Bacillus amyloliquefaciens contain a large open reading frame between the regions coding for signal sequence and mature protein. J. Bacteriol. 159, 811–819.

Vincentelli, R., Cimino, A., Geerlof, A., Kubo, A., Satou, Y., Cambillau, C., 2011. High-throughput protein expression screening and purification in *Escherichia coli*. Methods 55, 65–72, http://dx.doi.org/10.1016/j.ymeth.2011.08.010.

Wells, J.A., Estell, D.A., 1988. Subtilisin—an enzyme designed to be engineered. Trends Biochem. Sci. 13, 291–297, http://dx.doi.org/10.1016/0968-0004(88) 90121-1.

Zhang, R.H., Xiao, L., Peng, Y., Wang, H.Y., Bai, F., Zhang, Y., 2005. Gene expression and characteristics of a novel fibrinolytic enzyme (subtilisin DFE) in *Escherichia coli*. Lett. Appl. Microbiol. 41, 190–195, http://dx.doi.org/10.1111/j.1472-765X.2005.01715.x.

van den Ent, F., Löwe, J., 2006. RF cloning: a restriction-free method for inserting target genes into plasmids. J. Biochem. Biophys. Methods 67, 67–74, http://dx.doi.org/10.1016/j.jbbm.2005.12.008.

Zhu, X.L., Ohta, Y., Jordan, F., Inouye, M., 1989. Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. Nature 339, 483–484, http://dx.doi.org/10.1038/339483a0.